Université catholique de Louvain
Ecole Polytechnique de Louvain
Institute of Information and Communication
Technologies, Electronics and Applied Mathematics

Autonomous Frequency Containment Reserves from Energy Constrained Loads
A system perspective

Doctoral Dissertation presented by

Arnaud Latiers

in partial fulfilment of the requirements
for the Ph.D. degree in Engineering Sciences

Thesis committee:

Profs.	Emmanuel De Jaeger	(UCLouvain, advisor)
	François Glineur	(UCLouvain, advisor)
	Mathieu Van Vyve	(UCLouvain)
	Ronnie Belmans	(KULeuven)
	Göran Andersson	(ETH Zürich)
Dr.	Jonathan Sprooten	(Elia System Operator)

Louvain-la-Neuve, April 2016

Foreword

At the end of a four year-long journey, we put our last efforts in finalizing this text. I ought to thank the many guides that provided their help all along the way.

Firsts of them all, my two supervisors, professors Emmanuel De Jaeger and François Glineur. Their wisdom, patience, trust and attention where determinant to my success. Emmanuel's tremendous experience, *son lourd passé* in the electric industry was extremely valuable and kept our reasoning as realistic as possible. The impressive abstraction capabilities of François simply opened my eyes on problem solving issues. Also, as one may easily loose himself in the administrative Arcana, he provided the Ariadne's thread that allowed me to reach my final destination. Finally, their humbleness and open minds have considerably eased our discussions. I am really grateful to everything they brought to me, and their unconditional support, even more when deadlines were showing up a little earlier than I would have expected.

This work was conducted in close partnership with Elia, the Belgian electricity transmission system operator. In order of appearance, I firstly would like to thank Christophe Druet and Hubert Lemmens, that where the firsts to believe in this project. Their expertise pushed me forward and launched me in this adventurous journey in a great manner. Stephane Otjacques and Manuel Galvez rapidly followed, bringing their knowledge and support. Georgios Giannopoulos was of great help to me, especially considering his ability to share ideas and experience with its unique humorous skills. I also would like to thank Joeri Siborgs, the man who's always right. Thanks to Alexandre Torreele for bringing his open mind, trust, and great listening skills. He is a formidable challenger. Thanks to Arnaud Attanasi for his expertise and availability. I also would like to address a particular thank to Saskia Haillez and Els Wynant for their help. Also, I ought to thank all people from Elia that contributed to my work, particularly Jan, Peter-Jan, Emeline, Vanessa, Hans, Bob, Jean-Jacques, Melisa, Claire and Anna.

All of this would not have been possible without a great financing institution, Innoviris, and their idea of launching a university-industry partnership, Doctiris. It was a pleasure to discuss with them regularly about the state of our work. Xavier Lampe from ADRE contributed largely to our success.

I'm really grateful to all members of my thesis jury, that dedicated so many hours of their precious time for reading my work. I really appreciated to defend with all of them, some of which accompanied my PhD from the early beginning: Prof. R. Belmans, Prof. G. Andersson, Prof. Mathieu Vanvyve and Dr. J. Sprooten. Thanks to Prof. R. Jungers for presiding the jury. I am extremely impressed by the amount of comments I received, and the positive attitude that everyone showed during our four-hours-long discussion, without coffee.

A particular thank to Prof. G. Andersson and its team at EEH, ETH Zurich, with which I had the chance to work for a few months. In particular, many thanks to Theo, Olivier, Andreas, Stephan, Vaggelis, Marina and Michael (EKZ) for our interesting discussions. Thanks also to Osvaldito and Stavritos. For those who will know: *Let's proove our sons we were right.*

Many thanks to all my colleagues at CORE, CEREM and ICTEAM. Simon and Stephane greatly helped me during our brainstorming sessions. My co-authors, Julien and Benoit, for their scientific contribution and friendship, as well as Thomas and Baptiste should also receive my gratitude. Laurent Adrien and Camille were incredible lunch partners.

I started to get interested into this topic with my great friend Cedric Saussez. He helped me throughout this work to get my ideas clear, and he even corrected some chapters. Thanks for all, Ced !

During this project, I had the great chance to collaborate with several master students. They brought me a lot of inputs and our discussions were very interesting. Thank you for this, Jonathan and Cedric, Arnaud, Antoine, Louise and Emily. I'd like to thank Pierre Granville and Ludovic Honnay from Vivaqua as well as Xavier Yernaux from La brasserie du Bocq, with which we achieved very interesting projects.

I started with a background. I discovered electricity markets with the people of SEER, that will recognize themselves if they happen to read this text, and it was an incredible advantage.

To my parents, my two beloved sisters and their families : thank you for your support and patience. Many Thanks to all my friends, especially *les Feron-Cambier* and *Nutelet-Caudron*, to *the cousins*, to all the *in-laws*

(LGSAJ) and to Monsieur Cataldo, to the Altran's and associates, to la team de la sortie culturelle (C&K , A&G, P&A), to the flying's, and the gribouzes.

Last but not least, I am ever grateful to my beautiful wife Pauline. Her love stays, as a light glowing in the few darker times, and as sun rays underlying the happy ones.

Abstract

We are on the edge of a transition. Our fossil-fueled societies have to rapidly evolve toward decarbonated energy systems. Soon, we will rely on a more evanescent form of energy: the sun rays. As the morning comes and the night falls, this source of energy is intermittent, which is of particular concern to electric energy systems. Indeed, electrical power cannot be easily stored without being converted into a more convenient form of energy. Yet, the efficient storage of energy is a challenge that culminates in two extreme situations: when a large amount of energy must be stored either for long periods (seasonal storage) or in a brief amount of time (high power storage). Therefore, any alternative able to reduce the needs for energy storage infrastructures will soon both turn lucrative and vital to safe and efficient power system operations.

One of those alternatives is known as Demand-Side Management, which is a general term referring to all actions that can be undertaken to influence the amount and/or timing of electrical energy consumption. In a *decarbonated* system, relying principally on solar-based resources (Wind, Solar), Demand-Side management will aim at promoting energy consumption at the time energy is available.

On the one hand, moving energy consumption in time for a period of a month or more is hardly possible. *Who would wash its clothes in August to prepare for the winter time*? On the other hand, small demand adjustments performed during short time periods are technically feasible and much more acceptable on the consumer side. Furthermore, such adjustments can turn very valuable. Historical efforts in the area show that demand flexibility has shown particularly useful in several cases. One of them consists in preparing for major contingency situations which occur rarely but that could potentially have devastating impacts on the system. For instance, demand curtailments are often considered in case a sudden problem occurs in the aim of restoring the balance between the generated and consumed power. In certain cases, these actions are undertaken on a large number of specifically selected appliances: storage water heaters, refrigerators or even battery charging of electric vehicles. The law of large numbers makes such curtailments a very reliable resource.

However, the potential of exploiting such small appliances (electric loads) to support system operations on a more constant and frequent basis is still unclear. In theory, many loads can adapt their scheduled consumption with minimal impact on their end-user needs. But what if the ambition was to

extract as much flexibility potential as the loads can offer? For *how long* can you curtail the water heater *before* the user starts noticing ? As always, the devil lies in the details: how far can we push the loads implication into vital power systems operations such as real-time energy balancing?

In real-time, the power system operator takes very precise actions to maintain the system at equilibrium. When such precision is required from a large group of small electric loads, the control infrastructure responsible for shaping the group-level consumption is relatively advanced and complex. Unfortunately increased complexity leads in general to higher costs. This means that one cannot afford to control smaller appliances with great precision if those are not exploited at their full potential. Which brings us back to our initial problem: at the end of the day the consumer needs some minimal amount of energy to fulfill its needs.

> *In one way or another, electric appliances are faced with energy constraints as soon as their flexible capability is exploited at its full potential.*

This statement is one of the starting points of the present work. Our goal is to maximize the flexible power capacity offered by a group of small electric loads when those are participating in a real-time system balancing service called Frequency Containment Reserves (FCR). This service is an essential part of power system stability. It manages the system balance on a time horizon than spans between one second up to several minutes. The efforts deployed by the FCR providers consist in delivering large amount of power for short duration and at rare occasions. In consequence, such service is particularly well suited to energy constrained resources. Furthermore, FCR are generally a service of high value. The flexibility of small electric loads may also turn useful to other kind of services (e.g., peak shaving). Many of the theoretical results proposed in this work could be easily applied to those other context. However, our simulations will focus exclusively on FCR.

A second element is at the core of our work: system resilience. Resilience is the ability of a system to cope with change. The system components must have such intrinsic capability. The direct consequence is that the interactions between the small loads and the system as a whole should not get dependent on an additional infrastructure. The system must not rely on a dedicated communication infrastructure to control each participating load. This would indeed increase the risk of failures and decrease the system's resilience.

Loads can adapt autonomously (i.e., without external communication) based on two elements of information. The first comes from the system itself and is available at the plug, the connection point of the load with the network. It is the system frequency, one of the system's vital signs and an indicator of the current state of the generation-demand balance. The second is the knowledge that the load has about its own technical characteristics and that it can acquire about its user's habits.

> *What would be the overall system impact of large groups of energy constrained loads reacting autonomously to frequency changes for providing Frequency Containment Reserves ?*

To answer this question, we start from the simplest model one could think of: Energy Constrained Loads. This model represents a single electric load's consumption whose power is controllable but subject to energy constraints. Yet, such a model turns very enlightening for understanding the fundamental dynamics of the power demand of a large group of small electric appliances.

We then propose some autonomous controllers design able to act on the load's consumption and shape the group-level power demand in response to the system frequency changes. As a consequence of the above-mentioned energy limits, the provided response is not ideal. The load group is restricted to shift energy consumption in time, whose induces a specific kind of control errors called *energy rebound*. The rest of the electric system will have to compensate for these errors by using more of the slower flexibility resources it has at disposal. This has technical and economic consequences both in the short-term (seconds to minutes) as well as in the long-term (year long).

The long-term consequences can be grasped by conducting realistic power-system simulations spanning on months or years. Yet, it is impractical to simulate a large group of autonomously controlled loads on such long period of time. We therefore need to use *aggregate* models. These are simple mathematical structures that accurately represent the behavior of the group with reduced computational efforts.

Two families of aggregate models are proposed in this work that are relative to two general control frameworks. The first framework takes a simple approach in which loads react based on present information. The second is a more advanced case in which the past and its consequences (energy shifting, and consecutive energy rebound) are taken into account. As will be shown, the simple controllers lead to higher economic performances but cannot guarantee service in most critical situations.

Altogether, load control is economically efficient. Yet, the expected benefits in today's context are limited. In consequence, massive implementation programs and standardization seem to be the adapted processes to access the flexibility of small flexible electric loads. As the impact on the end-user is often imperceptible, including these controllers as a standard feature of most interesting appliances would be very well accepted. As regards to larger loads (commercial or transportation), benefits are such that even voluntary programs are affordable.

In 2016's context, it is counterproductive to target all types of energy constrained loads. The flexibility needs of the electricity system are not yet high enough to justify investing in control infrastructure for loads which yearly energy consumption is below a certain threshold. Addressing only the most interesting loads is key to ensure an overall positive societal impact. To give an idea, covering all FCR needs in continental Europe would require to control about 100 million loads or more. Certainly, this huge fish cannot be caught in one day.

Another major conclusion of this work is that the system operator's role is crucial in the overall profitability of the proposed control schemes. Its actions, and particularly the way it manages slower flexibility resources, have an important impact on the load control performances. Indeed, the technical and economic consequences of the energy rebound are reduced if the system can anticipate for it. Anticipatory actions both reduce the rebound itself, but also the costs associated with its management. As the system can forecast the rebound impact from monitoring the frequency deviations, it has time to prepare counteracting actions and call for low-cost and slower resources. Anticipating the immediate future has therefore a very high value in power system operations.

All questions were not answered in this work, but we could certainly highlight the fact that affordability also lies in the hands of the policy makers and system operators. As their interest into these issues is rising, small loads will most likely be part of the future flexibility mix.

Contents

Publications

J. Jomaux, A. Latiers, and E. De Jaeger. "Cost-based dimensioning of Battery Energy Storage and energy management system for Frequency Containment Reserves provision". In: *Power Energy Society General Meeting, 2015 IEEE*. Denver, July 2015, pp. 1–5.

A. Latiers, E. De Jaeger, and F. Glineur. "Toward decentralized primary control from demand response". In: *CIGRE 2014 belgium symposium*. Brussels, 2014.

A. Latiers, F. Glineur, and E. De Jaeger. "Energy limits in primary frequency control with short-term frequency-band allocation". In: *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*. Istanbul, Oct. 2014, pp. 1–6.

Arnaud Latiers. "Demand Response Perspectives for Belgium". In: *Reflets et perspectives de la vie économique* LIV.1-2 (2015), pp. 185–203. ISSN: 0034-2971. DOI: 10.3917/rpve.541.0185.

Arnaud Latiers, François Glineur, and Emmanuel De Jaeger. "Residential demand management and distribution grid impact assessment." In: *CIRED Lyon Conference*. 2015.

Arnaud Latiers, Benoit Martin, Francois Glineur, and Emmanuel De Jaeger. "On Decentralized Control of Small Loads and Energy Rebound within Primary Frequency Control". In: *Proceedings of the 19th Power Systems Computation Conference (PSCC)*. Genoa, 2016.

Benoit Martin, Emmanuel De Jaeger, Francois Glineur, and Arnaud Latiers. "A dynamic programming approach to multi-period planning of isolated microgrids". In: *Proceedings of ISESO International Symposium on Energy System Optimization*. Heidelberg Institute for Theoretical Studies, Nov. 2015.

List of Notations

Acronyms

AC	Air Conditioning
ACE	Area Control Error
aFRR	automatic Frequency Restoration Reserve
BRP	Balance Responsible Parties
DR	Demand Response
DSM	Demand-Side Management
DSO	Distribution System Operator
ECL	Energy Constrained Loads
EIA	Energy Information Administration
ENTSO-E	European Network of Transmission System Operators for Electricity
EU	European Union
EV	Electric Vehicle
FAPER	frequency adaptive, power-energy re-scheduler
FCR	Frequency Containment Reserve
FERC	Federal Energy Regulatory Commission
HVDC	High Voltage Direct Current
ICH	Interruptibility Contract Holder
I/D bids	Incremental/Decremental bids
IEA	International Energy Agency
IGCC	International Grid Control Cooperation
LFC	Load Frequency Area
mFRR	manual Frequency Restoration Reserve
NERC	North American Electric Reliability Corporation
R1	Primary Reserve
R2	Secondary Reserve
R3	Tertiary Reserve
R3DP	Tertiary Reserve Dynamic Profile
RR	Restoration Reserve
SO	System Operator
SOC	State-of-Charge
TCL	Thermostatically Controlled Load
TSO	Transmission System Operator
UCTE	Union for the Co-ordination of Transmission of Electricity

Symbols

f	System Frequency
f_n	Natural System Frequency (e.g., 50Hz)
ϕ	Normalization Frequency Band
$r(t)$	Normalized Frequency Deviation
$r^{up}(t)$	Normalized Frequency Deviation (upward)
K_D	Demand FCR Gain
K_{FCR}	FCR Gain
K_{aFRR}	FRR Gain
FRR_{EnR}	Additional Frequency Restoration Reserve
T_n	Natural Run Time of the load
P_n	Natural Power of the load
E_n	Natural Energy Need of the load
Ω_x	Probability Distribution of parameter x
T_{dl}	Time Deadline for Energy Delivery
T^{run}	Run Time of an individual load
T^{dyn}	Dynamic Run Time of loads in the Group
$D(t)$	Demand of the group
D_0	Baseline Demand of the group
v_D	Demand Volatility
$x(t)$	Group Demand Change
$N(t)$	Number of Running Loads
$S(t)$	Arrival process (starts)
$C(t)$	Counting process
λ	Arrival Rate

Miscellaneous

e.g.	For example (*exempli gratia*)
i.e.	That is (*id est*)

Units

Hz	Hertz
kW, MW, GW	kilowatt, Megawatt, Gigawatt
kWh, MWh, GWh	kilowatt-hour, Megawatt-hour, Gigawatt-hour
s, h	Second, hour

Chapter 1

Introduction

New York, 1891. In a crowded amphitheater, a modern alchemist performance fascinated the audience.

- *Sir Tesla, if I may, what is electricity ?*

Tesla paused and calmly looked at the journalist. He knew that he couldn't actually answer the question.

- *Ether under strain and Ether in motion are what characterize most probably electrostatic and -magnetic phenomena [1].*

Just about a decade earlier, the most ambitious entrepreneurs of their time started to build a structure of size and complexity that had never been seen in any previous mankind invention: the power system.

The power system is an infrastructure that allows an energy source to serve a distant load at the speed of light. It has a specific characteristic: as soon as energy is generated it must be consumed somewhere and cannot be stored on the network lines. The generated power must therefore adapt to demand variations. In practice, reaching the equilibrium between generation and demand constitutes one of the hardest tasks under the responsibility of the system operator. It generally requires to schedule precisely power generation to meet capricious demand levels. Along with the years and practice, the system has been operated with growing complexity. But there always was this persistent idea: why shouldn't electrical demand adapt to the available supply ?

1.1 Research question

This work explores the integration of small flexible electric appliances within short-term power system operations and frequency control. The focus is set on Energy Constrained Loads (ECLs), a general type of electric appliances characterized by a flexible power rate and predetermined energy needs over a given time horizon. The timing and power profile at which the load consumes energy are constrained by technical limits and user-related constraints.

We explore the ways of controlling the power consumption of small ECLs aggregated in large groups. In particular, the aim is to design autonomous controllers that are able to manage the group-level demand in a predetermined way while relying on a minimal amount of information. Autonomous control is a form of decentralized control setting in which no communication layer is considered. Loads must react only based on information they can obtain locally.

The purpose of this work is to explore the *overall system impacts* of autonomous ECL control delivering Frequency Containment Reserves (FCR), also known as primary frequency control. The FCR participants provide a flexible power capacity to the network operator that must be controlled precisely in response to the system frequency deviations they locally measure. Moreover, their response acts proportionally to the measured frequency deviations.

Loads must be controlled to react to frequency deviations while coping with their inherent energy constraints. These induce unavoidable control errors with negative consequences on the system frequency stability. Slower flexibility resources are therefore called more often to counteract these errors. This has technical and economic impacts that will be evaluated in both a short and long-term perspective. Thereby, we will be able to assess the profitability of investing in local controllers depending on the load characteristics. We also need to assess the impact on the end-user, and determine if these deserve additional payments for the service they offer. We must finally assess the consequences on power system operations. The system operator will most likely need to adapt its current operations to the specific behavior of the proposed load control.

If the overall economic impact turns positive, the fact that such a control scheme has not yet been implemented in the past must be explained by some other reasons. We will therefore make sure to identify all the barriers that may prevent the development of small load control.

1.2 Outline

Our research question is addressed in several chapters, whose content is described below. In Chapter 2, we look at the role of demand in power system operations. Our goal is to understand the historical developments in the area and assess its future potential. In particular, we will look at the situation in the Belgian system. In Chapter 3, the past and present use of small loads within frequency control is explored. In particular, practical and theoretical considerations on two types of load are scrutinized: Energy constrained Loads and Thermostatically Controlled Loads. This review concludes by identifying the research questions that have not yet been explored.

The consumption of large groups of ECL is studied in Chapter 4. The objective is to evaluate the intrinsic variability of the power demand of load groups based on the probability distribution of the load parameters and on the rate at which loads arrive in the group (starting time). The minimal group size that leads to sufficiently low demand variability is evaluated.

The main concern of Chapter 5 is to design autonomous controllers that can exploit the information extracted from the locally measured frequency deviation and make decisions on the load state while respecting local constraints. In order to analyze the impact of such simple control schemes in a long term perspective, the chapter ends by proposing aggregate models able to represent the group-level demand in a computationally efficient way. These models are derived for loads groups that share different parameters (heterogeneous groups).

An economic perspective is taken in Chapter 6. Are such controllers affordable? Is the overall impact on the system positive? These are the questions that will be answered in two types of power system operations: event-based (critical situation) and long-term historical simulations (one year).

In Chapter 7, we propose an advanced type of autonomous power controller that is able to remotely manage and shape the rebound error of the group. Based on the loads characteristics and constraints, we evaluate the amount of flexibility that can be offered as FCR while imposing the rebound error to follow a predetermined profile. The impact on loads and on the system is assessed, both from the technical and economic point of view.

Finally, Chapter 8 concludes our analyses.

Figure 1.1: Another summarized view, generated from `http://www.wordle.net/`.

Chapter 2

Demand Response in a nutshell

Chapter summary

From their early origins, many are those among the power systems worldwide that have tried to exploit the flexibility of electricity demand. In today's rush for flexibility, every and even the tiniest energy storage capability appears as a gold nugget to the neophyte. It seemed therefore necessary to investigate the successes and failures of past and present developments in the area. Let's hope that recalling the shoulders on which we stand will give us a somewhat more realistic flavor while daring to use the word "new".

This chapter is partly based on [89] from the author of this work, and published in 2015 in *Revue et Perspective de la vie économique*. Our goal is to give a very brief overview of the use of Demand Response among several systems worldwide, mainly in Europe, America's and Oceania. Our question: what explains the DR successful developments in some systems while others see almost no DR participation?

2.1 Power systems organization

According to the European Network of Transmission System Operators for Electricity (ENTSO-E) *the Power system comprises all generation, consumption and network installations interconnected through the network*[1]. The main goal of this system is to feed all the appliances (loads) that are connected to its network with a power of *good quality*. That is, the provided power must have predetermined characteristics such that it can be exploited and converted into usable energy (mechanical, thermal, etc.).

A process with a unique objective: safely meet the electric demand

The power entering an electricity network cannot be stored within its lines and cables. As soon as it is produced, the energy must flow through the network and reach its consumption site. This feature forces the operator generation assets to schedule their use according to the expected level of

[1] See Glossary at `http://emr.entsoe.eu/glossary/bin/view/GlossaryCode/GlossaryIndex`.

electricity demand.

The exact organizational process that defines how power generation is managed has evolved together with the socio-economic context in which the power system was anchored. Yet, three tasks must be performed invariably: (A) investing in infrastructures; (B) planning the use of these infrastructures; and (C) reacting to unexpected events or failures. In more complex terms, this leads to: (A) adequacy and security; (B) efficient planning; and (C) real-time balance. These tasks or goals are illustrated on figure 2.1 and separated into five *operational phases* (i.e., circled numbers one to five on Fig.2.1). While the first of these phases organizes the investment process, the four subsequent ones represent different time horizons for organizing the energy transfer from its generation site to its delivery point. They are detailed in the text below.

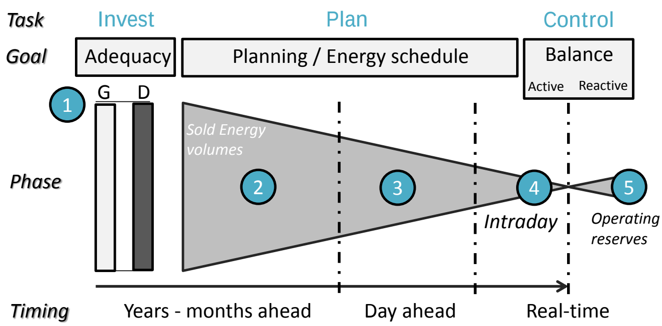


Figure 2.1: Power system operational phases (adapted from [106]).

1

Invest: Generation and Network Adequacy. An adequate system (G=D) has driven sufficient investments in generation capacity (G) to cover its highest power demand (D) level. The network is sufficiently strong to deliver energy where and when it is needed while coping with possible failures.

2-4

Plan: Energy generation and transmission schedule. Energy generation is scheduled in three successive stages. Large generators secure their energy sales years/months in advance (phase 2) to avoid uneconomic set-point adjustments. From a week to a day ahead of delivery (phase 3), smaller generators' schedule is decided. It may be refined *intraday* (phase 4) up to one hour before delivery. The network transfer capacity is allocated in a similar temporal sequence.

4-5

Balance: Energy balance and network security. On a near real-time basis, fast and precise control is vital to system stability. The system operator is responsible for coordinating the output of the many generation assets connected to its network. Ensuring the balance of generation and demand in real-time is the role of *operating reserves*. Reserves are composed by flexible power capacity excluded from the previous energy scheduling phases. Fast reserves enter into action right after an incident (e.g., loss of generation). Their burden is transferred to slower resources (e.g., slow reserves, market participants) in case failure is sustained for long periods. The balancing process stabilizes the system's vital sign: the system frequency.

In the first and tiny electrical networks, these three tasks were handled by the few system owners. With the system growth, multiple types of stakeholders with new roles and responsibilities. Their interactions, precisely defined in codes or laws, constitutes the core of the *design* of the market in which these actors evolve. We introduce briefly below the current design of the European electricity markets.

2.1.1 The core of EU system design: the balancing responsibility.

The first electricity networks had a very variable demand. The use of energy was very different from our today's habits. Power was available during rather short-periods of time (in the evening for lighting, etc.) [15]. Only progressively did the energy practice evolve towards increased consumption during off-peak periods. Together with this more constant use of energy, larger systems, interconnections, and better practice lead to decreased risks and higher power reliability. Yet, originally, there was a single captain: the system operator. The system balance responsibility (or constraint), that is the required equilibrium between the system generation and demand, was *centralized* in his hands.

The interconnected EU systems nowadays are market-based systems. The particular feature of the EU systems is that the power balance constraint is *decentralized* [56]. In short, every single network connection point (generation-side, demand-side) must be part of a portfolio under the responsibility of a specific actor called the Balancing responsible party (BRP). For every market period of a quarter hour, the energy produced by the BRP's generators or purchased from other BRPs should equal the sum of its energy sales (to other BRPs) and its clients consumption. BRPs are financially accountable for the net energy surplus/deficit of their portfolio.

The decentralized responsibility is grounded on the following principle: the market design gives all the incentives to BRPs for having a balanced system in real-time at the least possible cost. The individual efforts will balance the whole system. As some imbalance (i.e., surplus/deficit of energy) will always exist, the transmission system operator (TSO) will take care of the operations close to real-time.

A real-time market exists for settling the necessary transactions that were enforced by the TSO for reaching the final equilibrium. Long portfolios (BRPs that have too much energy) together with the activated reserves sell[2] their extra generated energy to BRPs with short positions. The real-time price of energy is the consequence of these interactions. If a spread exists between the real-time price and a BRPs portfolio marginal cost, he will get an incentive to react (modify generation/demand) for profit maximization. As the price indicates in theory the state of the market (in excess/deficit of energy), the system gets balanced from the decentralized BRPs reactions. The TSO operates both the transmission network (i.e., the high-voltage network) and the real-time (balancing) market. The zone under its responsibility is denoted as LFC area [52]. In principle, every LFC area dispose of its own real-time market.

In practice, the mechanism works well for standard and well predictable situations. As soon as price volatility increases, the message gets blurry: BRPs cannot decide from observing prices which actions to take. In addition, the classical pricing approach does not necessarily contain the correct information in real-time. Indeed, price levels are often influenced by both past and future (forecast) situations [4].

In opposition to the EU system, some market-based systems work with a centralized balance responsibility (i.e., pool-based systems). Independent generators are *pooled* together: they bid part of their capacity to a central entity that decides based on welfare maximization on their *economic* generation set-point. In real-time, the same principle applies to fast generation. This other way of organizing the generation schedule is also known as *central dispatch*.

2.1.2 The Supplier and the supply risk.

An other important role exists in the EU system: the Supplier. This actor is responsible for contracting electricity consumers and purchase sufficient

[2] All of theses transactions are essentially settled in the following week/month.

energy to cover their consumption. The supply contract essentially provides energy at a predefined *price* to the client. Each client in the supplier's portfolio is associated to one or several BRPs. Most often, Suppliers and BRPs form a unique entity. Yet, it is an additional layer separating the end-user from the core of the system.

The supplier's role introduces a distinction between two types of markets that co-exist in the EU design: the wholesale and retail markets. On the one hand, all energy trades occurring between large players form the *wholesale* markets. Multiple types of wholesale markets exists, organizing trade in different time horizons (see Fig. 2.1). On the other hand, the *retail* market consists in the contractual agreements between suppliers and customers.

Depending on the type of contract at stake, the end-user can be either completely isolated from or directly exposed to the prices of the wholesale markets. Getting isolated from the wholesale price fluctuations has a cost: the supplier adds a insurance premium to the average wholesale price. Therefore, consumers that have the capability to adapt their energy use will preferably choose contracts in which their contractual price is linked to the wholesale price.

In the wholesale markets, suppliers can buy energy in different manners. The two most common ways are either the self-scheduling (i.e. suppliers that also own generation assets produce for their own clients) and bilateral agreements (e.g., the supplier buys directly from a plant owner). The remaining volumes are traded on central auction markets. Such markets collect two types of bids: (1) the bids of generators which are the volumes they can provide at the associated costs; and (2) the bids of the demand-side representatives which define the maximum price they would pay for the energy volume they are asking. These bids are respectively piled up in ascending (generation) and descending (demand) order of price levels. The price at which the intersection of generation and demand curves meet decides on the price of energy, known as the clearing price. Such price applies to all energy volumes that are *cleared*. That is, all generation volumes whose bids is sufficiently low (i.e., below the clearing price) have been matched to a demand counterpart (i.e., which asked price is above the clearing price). All other demand and generation bids are rejected.

There exists therefore a risk for the supplier: it could have bought an inadequate amount of energy. In real-time, though, the system must be balanced. Some extra transactions must take place such that, at the end of the day, each supplier has bought the exact amount of energy that was

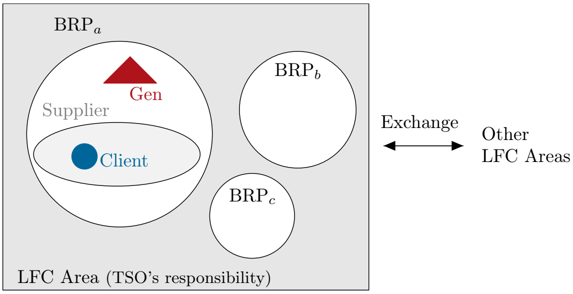


Figure 2.2: Role model of the EU electricity systems within each synchronous area (group of LFC areas).

consumed by its clients. This is where the BRP role enters into force. In some ways, the notion of BRP serves a risk management purpose for suppliers of small size.

2.1.3 Summary view of the EU role model

All the above described roles are summarized on figure 2.2. The precise definition of the roles and responsibilities of the different actors along with the way they can interact with each other constitutes the *market design*.

As highlighted on the figure, some energy exchange between neighboring LFC area will emerge from international trade. In real-time, the TSO and the balancing market will make sure that the exchanged power corresponds to its scheduled level.

2.1.4 Ancillary services for frequency control and the links with the balancing market.

Ancillary services is a general term that encompasses different safety and reliability measures at disposal of the grid/system operator. These services are dedicated to frequency control, voltage control and to system restoration after a black-out (black start service).

The frequency control services are the operating reserves which are described in the recently published ENTSO-e Network codes [52]. Their role is to maintain the system frequency close to its nominal value (i.e. 50Hz in EU systems). Fig. 2.3 shows the sequential approach of the frequency control scheme.

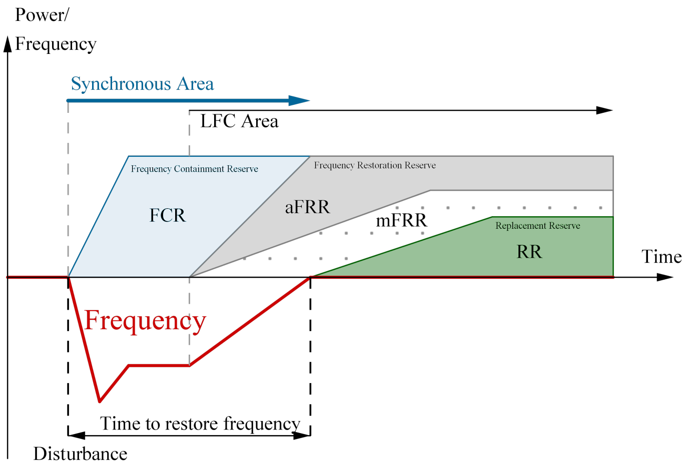


Figure 2.3: Entsoe load-frequency control scheduling process [52].

The Operating Reserves consist in an insurance product. The System Operator wants to guarantee ahead of real-time that it will have sufficient flexible capacity at disposal to balance the system. The deployment of the reserve capacity follows a step-wise replacement scheme (see Fig 2.3) and a spacial hierarchy.

Damping frequency changes: Frequency Containment Reserves

> *Frequency Containment Reserves (FCR), also known as primary reserve, are the Operating Reserves activated to contain System Frequency after the occurrence of an imbalance.* [52]

Following an incident (e.g., sudden loss of a generator) a first frequency stabilization process *automatically* takes place: Frequency Containment Reserves (FCR). It is synchronously performed across the whole system (Synchronous area). Each LFC area that is part of the synchronous area contributes to this system-level control in proportions of its size. In practice, a set of fast reacting assets must adapt their power generation (or consumption) in proportion of the measured frequency deviation from nominal. The FCR reaction is therefore zero when the frequency is at nominal. FCR provide some damping effect to frequency excursions and lead these to stabilize at a steady-state value.

Equation (2.1) governs the evolution of the system frequency $f(t)$ with time t, and more particularly its deviation $\Delta f = f - f_n$ from its nominal value f_n (without FCR contribution).

$$M\frac{d\Delta f(t)}{dt} + L\Delta f(t) = \Delta P_g(t) - \Delta P_l(t) \tag{2.1}$$

The frequency change $\Delta f(t)$ is influenced by the change in generation ΔP_g and/or load ΔP_l w.r.t. their value at $f = f_n$ (i.e., the imbalance) [111]. After an imbalance has occurred, the lack/surplus of power is compensated by a change in the kinetic energy of the rotating generators (direct coupling) which modifies their angular frequency or rotational speed. The consecutive frequency changes are slowed down by the angular momentum ($2\pi M$) of these rotating masses that depends on their total inertia I (M=If_n). This inertial behavior acts as a low-pass filter on the initial imbalance. In addition, a proportional damping effect L opposes to further frequency deviation. This effect originates from the natural frequency dependence of some electric loads and of the generators winding.

The introduction of FCR in equation (2.1) will further limit the frequency deviations and *contain* the frequency within a small interval around its nominal value. FCR are built to provide an additional damping effect of gain K_{FCR} and proportional to the observed deviation.

$$M\frac{d\Delta f(t)}{dt} + L\Delta f(t) = \Delta P_g(t) - \Delta P_l(t) - K_{FCR}\Delta f(t) \tag{2.2}$$

Steady-state conditions are reached for $d\Delta f(t)/dt = 0$. In the simple case for which ΔP_g and ΔP_l are constant, we have the following steady-state deviation (eq. (2.3)).

$$\Delta f = -\frac{\Delta P_g - \Delta P_l}{(L + K_{FCR})} \tag{2.3}$$

The ENTSO-e Network Code on Load-Frequency Control [52] defines a reference incident corresponding to a N-2 security criterion. In Continental Europe (CE), the N-2 situation is the sudden loss of the two largest generators: two 1500 MW nuclear reactors. This reference defines the minimal FCR proportional gain K_{FCR} [MW/Hz], also known as *Droop* or K-factor, that contains frequency deviations within desired limits. The frequency deviations should stay below the *maximum steady-state frequency deviation* Δf_S (= 200mHz in CE).

$$|\Delta f| = \frac{3000MW}{(L + K_{FCR})} \leq \Delta f_S = 200mHz \tag{2.4}$$

The K-factor should be sized such as to maintain the frequency below its maximum steady-state value while neglecting the contribution of frequency-dependent loads and generators winding (i.e. $L = 0$) [145].

$$K_{FCR} \geq \frac{3000MW}{200mHz} = 15000MW/Hz \tag{2.5}$$

Each LFC area participates to the synchronous area K-factor in proportion of its size. Let's finally note that the deployed FCR is defined positive for generation increase. It is also divided into upward (generation increase) and downward (generation decrease) volumes.

$$FCR(t) = -K_{FCR}\Delta f(t) \tag{2.6}$$

Recover nominal frequency: Automatic Frequency Restoration Reserves

> *Frequency Restoration Reserves (FRR) means the Active Power Reserves activated to restore System Frequency to the Nominal Frequency and for Synchronous Area consisting of more than one LFC Area power balance to the scheduled value.*[52].

In a second step, each LFC area takes responsibility for its own internal balance. Restoring every local balance eventually leads the whole system to get balanced and simultaneously brings the frequency back to its nominal value. The FCR participating resources return their initial set-point and are ready to counteract another incident. The local balance can be restored from calling two other types of reserved capacity: (1) automatic frequency restoration reserve (aFRR); and (2) replacement reserves (RR).

The automatically activated FRR are controlled to oppose a composite error known as the Area Control Error (ACE).

$$ACE(t) = \Delta T(t) + K_{FRR}\Delta f(t) \tag{2.7}$$

$\Delta T = T(t) - T_0$	Difference between measured $T(t)$ and scheduled T_0 power flows at the area border (> 0 increased export).
K_{FRR}	The frequency bias. It sums the area's damping factor L with its contribution to the synchronous area's FCR K-factor.

Let's imagine a synchronous area composed of two LFC areas (1 and 2). A sudden imbalance occurs in area 1. The change in external flows $\Delta T(t)$ corresponds to the FCR contributions of the other zone, with opposite sign due to the flow direction definition. The ACE computed in zone 1 is the sum, with opposite sign, of its own contribution to FCR $K_{FRR}\Delta f(t)$ with the one of the other area's, and therefore corresponds to the initial imbalance. On the other hand, the other area has an ACE equal to zero. Therefore, the ACE definition (2.7) lead each area to counteract its own shares of the overall synchronous area imbalance. In practical situations, this works well provided that the K_{FRR} factor represents the actual frequency reaction of each LFC area.

Compensate for sustained imbalances: RR and the balancing market.

In case of sustained disturbance, aFRR will get (partially) relieved by slower resources originating either from (1) slower operating reserves (Replacement reserve or RR), and/or (2) all other participants to the *balancing market* (real-time market). RR are actually the less frequently called resources. Indeed, their main purpose is to be available to cover the largest imbalances that may occur.

Therefore, the goal of the balancing market in normal operating conditions can be understood as a way to *select* the most efficient resource able to free-up the automatically activated (aFRR) volumes. Several options exist.

- Non-reserved flexible resources may participate to local balancing. In opposition to reserve capacity, these participants do not need to guarantee the provided volumes in advance. Instead, any resource having unused capacity at disposal is *forced* to offer it at a *freely decided* price. The participants can be demand-side (i.e. large customers) or generation-side assets. In Belgium, these offers are called *free-bids* or I/D bids (incremental/decremental). After aFRR, I/D bids are the second most often activated resource in the Belgian system.

- A solidarity mechanism exists between neighboring LFC areas. This mechanism introduces a link between their respective balancing markets and real-time price. It is called IGCC (International Grid Control Cooperation) also known as ACE netting. *It consists in the exchange of opposite imbalances* (ACE of opposite signs) *in order to reduce the imbalance that has to be compensated* by *each control area*[3]. Part of the interconnection capacity (tie-line) must be made available for this specific purpose.

In Belgium, the pricing mechanism at stake is called *single marginal pricing*. The real-time price is defined uniquely for each quarter hour. In

[3]`http://www.elia.be/en/products-and-services/balance/balancing-mechanism`

case of net lack (resp. surplus) of energy, the price corresponds to the cost of the most (resp. less) expensive resource (i.e., aFRR, RR, IGCC, I/D bids) activated by the TSO. The real-time price serves as an indicator of the system state and triggers an indirect reaction from the different BRPs.

In the Belgian system, FCR are called primary reserve (R1). The aFRR are secondary reserve (R2) and the RR are declined in three products: tertiary reserve (R3), tertiary reserve *dynamic profile* (R3DP) addressing resources that are not directly connected to the TSO's network (i.e. lower voltage levels) and ICH which corresponds to curtailing large industrial consumers.

A typical day in Belgium

A typical daily activation profile of the different resources is shown on figure 2.4. Three important elements should be noted.

- FCR are rarely deployed up to their available capacity as sized to cope with critical events.
- The manually activated resources are slow. Prior to their activation, the TSO must assess the chances for aFRR to reach their limit ($\simeq \pm 140$MW). In the Belgium, this task is extremely difficult due to the relatively small size of the system. Therefore, saturation of the aFRR capacity is not rare. The relative energy content (integral) of aFRR is much higher than the one of FCR. Therefore, limited-energy resources such as storage are not the best candidate for providing aFRR volumes in Belgium.
- The volumes colored in red on figure 2.4 are the I/D bids (no RR activation, as often the case).

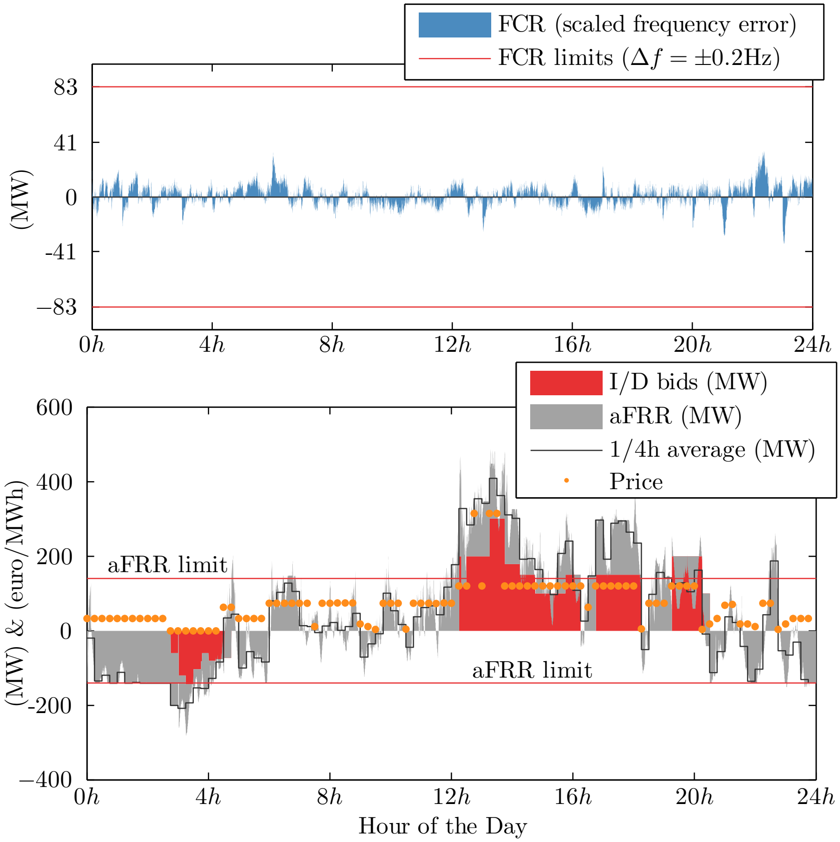


Figure 2.4: Activated reserve volumes (10-second average) in the Belgian LFC area on the 5^{th} of January 2014. Top chart: FCR volumes (limited by ±83MW deployed at $\Delta f = \mp 0.2$Hz). Bottom chart: aFRR and manually activated volumes for area balancing and frequency restoration (aFRR limit of ±140MW). The real-time prices is also shown (dotted), while IGCC volumes are omitted for readability. Source: Elia System Operator.

2.2 The Role of demand in system operations.

The *traditional* system operations enforce a few flexible generators to adjust their power output and follow demand fluctuations while slower generation adapts to cover long-term demand trends[4]. Many Demand-side assets, from large industrial processes to tiny household appliances, adapt their behavior to help the system accommodating changes. The set of actions undertaken to influence electrical demand are referred to as Demand-Side Management (DSM). This section explores the past and future role of demand in system operations.

2.2.1 A renewed interest for demand-side management: Future Flexibility needs of Belgian's electricity supply

Among the possible ways in which Belgium's energy supply could evolve, electricity is often perceived as taking an increased share in the fuel mix. Solar-based power is considered able to limit fossil-fuel dependency. Furthermore, converting appliances to electricity (for example, electric transportation, heat pumps) could generally make the use of energy more efficient. However, the solar-based generation requires storage infrastructure that can efficiently limit the impact of their variability across hours, days, and seasons.

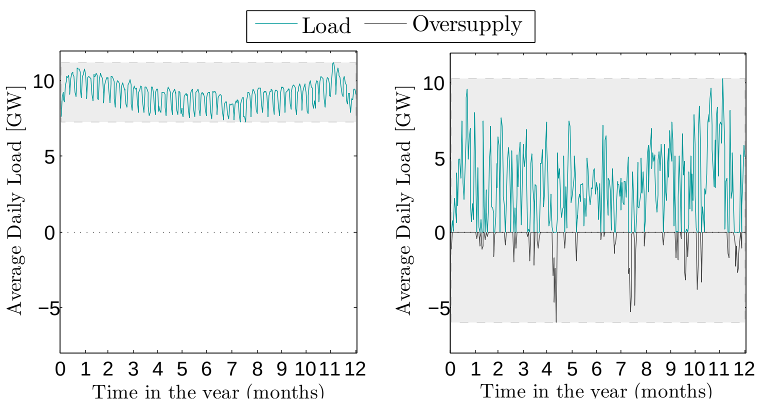


Figure 2.5: Evolution of the Belgian System load (average daily load in GW). Left: traditional load (2014). Right: Distant future net load (2050).

[4]Belgian Load variations: `http://www.elia.be/en/grid-data/Load-and-Load-Forecasts/total-load`

Figure 2.5 illustrates the daily *net* consumption of electrical energy in Belgium, that is the actual energy demand to which solar-based generation is (Wind, PV) subtracted. The left-hand chart shows the situation in (2014) while the right chart shows a possible future[5] (2050). In 2014, the daily energy demand, expressed as an average load, varied between 7 and 11 GW. In 2015, the power system must absorbs a much larger variability in the range -6 GW (energy oversupply) to 10 GW, which is close to today's maximum. In 2050, the non-variable part of generation (e.g., power stations, stored energy, energy imports, etc.) will need to accommodate high consumption days in a similar fashion to 2014's but will experience days with energy surplus, where all generation infrastructure should be shut down.

A massive amount of investment is needed to cover such variability range. One solution consists in exploiting extremely flexible back-up generation to cover the whole load range. Another resides in storing all of the oversupplied energy. However, the way forward lies between these two extremes. In particular, adapting the initial demand profile (demand-side management, DSM) is one option for alleviating the burden placed on the power system. DSM alone cannot fulfill 2050's needs for flexibility. Yet, it is definitely about to take on increased importance.

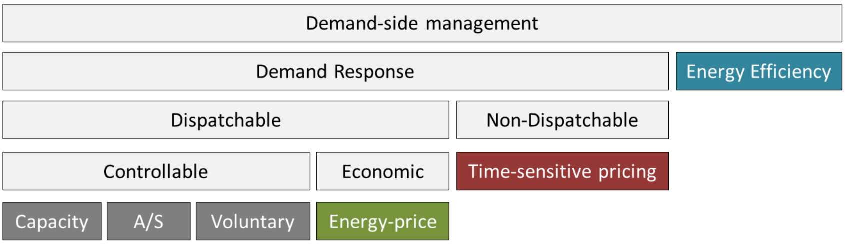
2.2.2 Definition of Demand-side Management

Demand-side management (DSM) encompasses two different concepts: energy efficiency and demand response. The first promotes long-term demand change (e.g. a new fridge consuming less energy) while the second consists in more dynamic and short-term demand change following external solicitations. The definition of the NERC is illustrated on Fig.2.6.

Demand Response (DR) consists in the set of short-term actions applied to electricity demand *in reaction to* critical system conditions (e.g., power shortages, contingencies, structural lack of energy). The existing forms of Demand Response can be distinguished by the degree of control they allow and the needs they fulfill: (1) capacity needs such as lack of generation to cover peak demand, (2) energy needs during high energy price periods or (3) ancillary services. There exist two different approaches to capture the response of electricity consumers: price-based and incentive-based programs. Each of these implementation schemes is adapted to particular consumer segments and to the specific service that is being delivered.

[5]Simplified CORE scenario from [33]: 2014 load, wind profile (×16), PV (×7.5) and CCS.

Figure 2.6: NERC classification of Demand-Side management [114].



Demand-side management	All activities or programs undertaken to influence the amount and timing of electricity use.
Energy efficiency	Permanent changes to electricity (e.g., more efficient end-use devices). Generally, it results reduced consumption across all hours rather than event-driven targeted load reductions.
Demand Response	Changes in electric use by demand-side resources from their *normal* consumption patterns *in response* to changes in the price of electricity, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized.

Price-based DR: Pricing schemes for indirect response.

Price-based DR are based on energy contracts in which the consumer is exposed more or less directly to the wholesale energy price (day-ahead, real-time). Price signals trigger indirect response from the consumer. Several options exist that express different degrees of complexity and levels of risk-sharing between electricity suppliers and their clients [58]: from invariable peak/off-peak tariffs to full integration of consumers into wholesale markets (Economic DR) or even real-time markets [82]. The most complex contracts are suited to large consumers that have a precise control on their electric appliances or industrial processes.

Theoretically, price-based DR are the most attractive programs. In theory indeed, a consumer can limit its consumption during high price periods if costs rise above its own willingness to pay for energy [27]. Furthermore, it decreases the *price* risk of supply: the risk of seeing wholesale prices rising above the retail prices (contract-based) is lower when retail prices are adaptable. The interested reader can see the toy example presented in Appendix A. Finally, the price-based DR has a retroactive effect on the electricity prices: high peak prices lead to decreased demand what eventually decreases the prices. This can be beneficial to society as a whole.

In the long-run, with appliance automation and measurements from advanced metering infrastructure (i.e., *smart* meter), price-responsive demand could become a successful tool for demand forecasting. In the short-run however, the introduction of price-based DR leads to increased risks and volatility as the reaction to price cannot be precisely anticipated. A system with significant participation of price-based DR can be hard to stabilize [4] especially if smoothed feedback is not implemented within the price formation [28]. Indeed, incorrect estimation of the Demand responsiveness leads to over/under-reaction and increased price volatility.

Tariff design is in itself a very complex process. As international experience has shown (see SEDC pilot review [134]), tariff design requires very good market segmentation (customer based studies), education and feedback (information to and from consumers about real-time consumption and energy consumption trend) and integration of appliances automation. Also, participating appliances should be selected with care, as the overall environmental impact of the installed controller is not always positive [137].

The most complex price-based DR never had a large success among smaller consumers [21]. Indeed, smart-meter alone are not sufficient. Their measurements must be reported for billing in a secure and verifiable way. A complete data processing procedure is required, sometimes too complex to be justifiable at the household level. Using smart-meters at their full potential requires large efforts in marketing and information, and suppliers may never have enough time and resources to achieve this [155].

Dispatchable DR: Triggering a direct response.

As an alternative to price-based DR, incentive-based DR programs *pay* the consumer for the amount of delivered flexibility. This latter has to be carefully determined by comparing the actual user's consumption to an hypothetical *baseline* consumption. Baselines are supposed to represent the user's power demand as it would have been *without* DR participation. It is in theory impossible to compute a correct baseline if the user did not precise it beforehand (nomination, purchase level). Indeed, users should *purchase* their baseline before being allowed to *sell it back* to the market as flexibility [124]. This is ideal, in a pure economic sense, as the consumer has no incentive to retain information from the market (strategic demand bid). It is however unpractical for small consumers subject to variable consumption.

Choosing between price and incentive based DR and adapting market/-contractual rules accordingly represents one of the most difficult tasks of program designers [67]. Typically, incentive-based DR is used in extreme case scenarios. Indeed, as the probability of DR action is low, the consumer has no incentive to prepare for it. Therefore, baseline computations are not biased and may be computed by regression on a large number of consumption samples (past consumption). This is the case of most incentive-based programs which goal is to solve adequacy issues or provide slow reserve. Also, some programs simply do not verify activation. For example, DR programs for network security consist in instantaneous and rare load shedding to support the system in case of large element failure (defense procedure).

2.2.3 Why and where is DSM developed? Find the sweet spot !

Demand-side management and more specifically Demand Response (DR) has been successfully deployed across the different operational phases mentioned above. A specific system always promotes DSM where it is the most critically needed. Every system has peculiarities creating a value pocket for flexibility, in other words, a *sweet spot* for DSM deployment.

The most fundamental motivations to deploy demand response is when system's survival is at stake. Structural system needs can be of different nature: (1) security issues (2) energy/capacity shortage (3) transfer capacity (4) asset aging.

- *System Security.* Systems with large elements (tie-line, nuclear power plant) face large risks in case of failure or unforeseen events.

- *Adequacy.* System whose generation mix is subject to hourly (power adequacy) or seasonal (energy adequacy) variations or systems with peaky demand profiles (temperature sensitive).

- *Network transfer capacity.* Systems where limited transfer capability induce local excess of generation (renewable energy, must-run generation) or load (cities, industry).

- *Asset aging.* Systems whose components are getting older may defer investments or reduce maintenance costs using DR to decrease local peak loads.

Load curtailments are commonly used in extreme situations. Their curtailment occurs rapidly which is very beneficial in emergency situations.

Furthermore, curtailment of a large number of loads is statistically more reliable (not more precise) than shutting down a few generators [28].

> *In some respects providing* [replacement] *reserves with demand response is similar to meeting peak loads with demand response. In both cases load reductions of a few hours per year are likely to meet the system need.* [117]

This quote illustrates that there exists some value pocket for demand in providing rarely occurring but high-value services, be it peak load decrease or security related services. DR is most often restricted to provide high value flexibility services (extreme cases) and/or services with relatively low precision requirements (curtailment). More frequent or precise control on loads requires to invest in adapted control/communication infrastructures which can be costly.

DR programs for enhanced System Security

System of peculiar topology or disposing of large demand/generation/transmission assets are more concerned with security issues.

New-Zealand is a typical example of system with peculiar topology. The two main islands of New-Zealand are connected since 1965 by a strong high-voltage direct current (HVDC) cable [143]. The northern island relies on the southern hydraulic energy and imports massive amount of power through the HVDC cable, particularly at winter time. In case of a sudden failure of the interconnection, the system frequencies and local voltages in the north and south parts may experience very large variations, possibly damaging generation and transmission infrastructures or implying cascading outages resulting in a blackout. To prevent this to happen, two layers of demand control are exploited: *interruptible loads* and *automatic under-frequency load shedding*. These gather loads of different sizes across all sectors, from residential to industrial [142].

The Alberta electricity system operator (AESO) runs a pool based (central dispatch) market and has developed a large number of products for price and system stability. All of these product are load curtailment products. Two products were developed for system security purpose [49] and aimed at increasing the import capability from the British Columbia inter-tie: Interruptible Load-Remedial Action Scheme *IL-RAS* and *59.5Hz Load Tripping*. Both product serve the same purpose than in the New-Zealand example.

Norway is the country that has the largest shares of hydropower worldwide (96% in 2013[6]). Hydro-power is usually the main provider of operating reserves. During Norwegian winter, resource become scarce. Therefore, a large part of the required reserve are provided by demand [69].

DR programs serving System Adequacy

Adequacy problems arise when generation experience seasonal variations (e.g., solar-based, hydro power) or when weather sensitive demand (e.g., electric heating, Air-conditioning) takes large shares of the system peak demand.

In California, 30% of the peak load comes from small air-conditioning loads [26]. The residential demand contributes for 70% of the peak demand in Texas. Direct load control of residential A/C has thereby been undertaken by retail companies, in close collaboration with the grid operators. In such systems, rather complex control programs are cost-effective.

In France, winter peak demand grew massively as a consequence of the massive adoption of electric heating, following the development of nuclear energy. The large french nuclear fleet providing 73% of the produced electricity (2013, IEA[7]) is not designed to follow the large seasonal variations but rather to provide a firm base generation level. A price-based DR programs was launched as early as 1983: the EJP option (now Tempo Option). The energy is charged differently according to the time of day (peak and off-peak hours) as well as the day in the year (blue, white and red days). Decentralized generation (behind the meter) was also allowed to participate as only the net consumption was measured and charged. In recent years, the effective demand decrease originating from those programs has felt dramatically. In 2000, the contribution of professional consumers was 4000MW and felt to 1000MW in 2007. As mentioned in [121], half of this sudden decrease has to be attributed to the decommissioning of decentralized generation (consequence of stricter environmental norms).

Most systems were a dominant part of the generation comes from hydro power are faced with adequacy problems. Systems with increasing shares of variable generation (solar, wind) are likely to face similar issues. The seasonality of hydroelectric generation, especially run-of-river, implies that there exists some days in the year were there is simply not enough water to power all demand [53]. This energy limit implies capacity shortage as

[6]BP statistical review: `http://www.bp.com/en/global/corporate/energy-economics.html`

[7]`http://www.iea.org/statistics/statisticssearch/report/?year=2013&country=FRANCE&product=ElectricityandHeat`

the available energy must be spread over the day. These capacity shortage either occur on the energy market (i.e., the peak load cannot be served) or on the reserve market (i.e., spare capacity to be used in case of contingency becomes scarce). Demand Response is especially well developed in such systems [42, 129, 116].

In Colombia, the hydro-power shares in the generated energy are about 71%[6]. The capacity shortages resulting from the recurrent storm *El Niño* forces the country to deploy specific tools for insuring adequacy. The Colombian market imposes firm capacity requirements, a capacity auction, in which demand-side resources are allowed to participate [101].

Hydro-power does not always engender *direct* adequacy problems. In the US, Washington State electrical generation comes at 77% from hydro power. The large number of hydro reservoirs (dams) experience low variability, creating an excess (about 24% above peak load) of capacity in the region. Energy prices are the lowest from all US states. In 2009, there was no demand response program in the Washington State [139]. In years of severe drought, dams may not gather enough water to serve all the energy demand [42]. However, the state is a net exporter of energy and can reduce its exportation level during these years. The neighboring States are the one that will be the most impacted. Yet, the growing peak demand, increased shares of wind power but also internal network constraints are changing the picture. Bonneville Power Administration has activities in Idaho, Washington and Oregon. It started DR programs in the agricultural (2010, irrigation), residential (2011, Thermal Storage), commercial/industrial sector (2011,Water pumping).

2.2.4 DSM and market design: An historical perspective.

The role that DSM takes in system operations worldwide is mostly a consequence of the specific physical characteristics of the system in which DSM takes place. Yet, market design plays also an important role in the degree of development of DSM programs.

The origins of Demand-side management.

At the end of the 19^{th} century, an interesting idea was developed separately by two great scientists and business men: Samuel Insull and John Hopkinson [133, 124]. Their electricity network had originally been built in order to power light bulbs, in great competition with the previously existing energy source for light: the gas networks. Similarly to today's context, the specific nature of electricity required production to adapt to consumption at all time. Technical requirements were largely under what is needed in nowadays large scale grids (i.e. small size, radial architecture, simple loads). Nevertheless, Insull and Hopkinson both had the profound intuition that consumers were to be more actively integrated in their network operations. Their idea consisted in the primitive form of least-cost planning and demand-side management. They wanted to optimize the production cost of electricity by acting on its demand. They immediately thought that an efficient way to involve consumers was to charge them for their consumption pattern rather than for their total energy consumption. Wright tariffs (S. Insull) or time-of-use pricing (J. Hopkinson) were two similar ways to foster consumer's response in the hope of improving the system's cost-efficiency.

The Wright tariffs were originally conceived by Arthur Wright, who invented the real-time meter [133]. Its basic principle was to charge consumption proportionally to the duration of each of its observed level (*load duration curve*) [31]. An end-user consuming 10 kW for 12h and 20kW for another 12h would be billed as such: (10kW$\times$24h)BR + (10kW$\times$12h)PR, where BR and PR correspond respectively to base and peak production costs. This tariff model assumes that the consumption profile of each individual user and the total system load profile are similar.

Nowadays, many dynamic tariff schemes can be found in the different systems worldwide. Their structure varies as they are designed to address system-specific issues (e.g., high peak load, variable generation). Indirect control programs (timers) were implemented on residential loads as early as 1934 [68].

A push from the oil crisis.

In the early 1970s, the first oil crisis led to a concern that electricity demand would not keep on its previously growth rates. Electric utilities reached a turning point: new investments and infrastructure renewal could not be justified by future demand growth. In this context, Amory Lovins compared two evolution scenarios for the electricity sector: demand conservation vs infrastructure growth. In the following decade, his conclusions had a great influence on decision makers. The first large-scale demand conservation programs were launched in the US. Demand-side management experienced a regain of interest. US congress firstly mentioned the term Demand-side management in the National Energy Conservation Policy Act (1978) [55]. Many new dynamic tariffs and direct demand control programs (mainly designed for Air Conditioning control) were implemented [124].

However, electricity rates were rising. Indeed, the regulated price had to be raised in order to compensate for the freshly endorsed investments. In the 1980s, following the oil price crash and gas market deregulation, electricity became less competitive. This led to a further decrease in demand. In the US, the regulation regimes and infrastructure differ largely from one state to another. Some states were therefore much more impacted than others: they faced both an overcapacity and high prices. Furthermore, DSM measures were largely subsidized as they resulted from analysis from the previous decade [157].

Liberalization and the (small) fall of DR programs.

This situation was unacceptable for the largest electricity consumers. Large industrial players urged for lower prices. At the same time, many economists had been interested into the electricity price formation mechanisms, considering electricity as a commodity rather than a public goods [14, 74, 20, 30]. The liberalization process was launched.

A the end of the 1990s, this had two main consequences. First of all, wholesale prices indeed fell in many systems. If deregulation most probably played a role, the effect of low gas prices seems to have been a dominant driver of this trend [21]. But the effective decoupling between the supply and generation function had large consequences on the demand-side programs initiated close to 20 years earlier. The newly created suppliers seen no interest in the subsidized DSM programs. Having sufficient capacity and flexibility was of no concern in systems with overcapacity. Consequently, a large number of DSM programs were stopped [157].

This is illustrated by the case of Texas electricity system, nowadays ruled by ERCOT.

> *Prior to the introduction of retail competition in January 2002, ERCOT relied upon roughly 3500 MW of interruptible load and other load management programs to maintain reliability [...] Innovative pricing programs had also proven successful.* [157]

The newly established market will suddenly lose these precious resources. The direct effect was a rise in operational costs and a decrease of the security of supply.

> *On January 1, 2002, the market lost a planning reserve resource of nearly 3000MW. Also lost was the under-frequency response from large industrial loads on instantaneous interruptible tariffs which was used to offset spinning reserves requirements under the previous utility structure.*

This happened because of the conflicting interests that emerged between the newly established players, the increased competition and the redefinition of the tariffs.

> *Restructuring required the termination of all tariffs in the areas of ERCOT opened to retail competition, including the tariffs offering a discounted price to interruptible loads and the tariffs used to provide consumers with real-time pricing options.* [157]

A similar dynamic was observed in 1997 in Belgium, when deregulation of electricity was implemented for industrial players. The Belgian system witnessed a waiver of interruptibility/modulation tariffs that were previously charged to numbers of industrial players. Two elements can explain this (small) decrease. First, the competitive context led the retailers to aggressively capture (or keep) as much consumers as possible, by proposing more favorable and/or more simple rates. These newly created players saw no direct interest in keeping the old complex contractual terms. Secondly, the smaller industrial sites got physically disconnected from the entity responsible for the system balance. After deregulation, some sites connected to lower voltage levels were attributed entered the portfolio of distribution grid operator (DSO). However, the Belgian technical rules were not adapted to legally exploit for system balance purpose the flexibility of clients that were not directly connected to the transmission grid operator (TSO).

Two years after the deregulation was launched, in 1999, the Belgian state[8] ordered a parliamentary commission to study all possible solutions for the future of electric supply[9]. The goal was to prepare for the expected decommissioning of the two nuclear power stations that are still running today and fulfill about 50% of the Belgian electricity supply. In the report, appeared the first official mention of the notion of demand-side management at the state level. However, the results of the commission's work has not received, at the time, the attention it deserves [100].

We can see that one of the main barriers to the implementation of DR programs is that the responsibility for implementation is not always clearly identified [77]. The different conflicting interests are usually slowing down the process of DR development. For instance, Suppliers perceive DR rather as a marketing tool than as an economic opportunity. Indeed, customer retention is the principal motivation for developing price-based DR programs[130] in the US. Such difference in perception is particularly damaging

[8] 25 NOVEMBRE 1999 - Arrêté royal portant modification de l'arrêté royal du 19 avril 1999 instituant une Commission pour l'Analyse des Modes de Production de l'électricité et de Redéploiement des Energies (AMPERE).

[9] Very interesting report : `http://arp83.free.fr/rapport_ampere.pdf`

when roles and responsibilities have just been created, as was observed in certain systems right after deregulation.

Counteracting Market Power and Investment risk: the new DR golden age.

The deregulation process (liberalization) had to face two main problems in the US: the missing money problem, and the market power problem.

1. *Missing money*: The newly created markets induced a riskier environment for investing into peak generation (i.e. generation assets that are used only at peak time). Such trend was highlighted notably by Cramton and Stoft and referred to as *the missing money problem* [36].

2. *Market Power*: The US regulation is divided into a state and a federal level. In some cases, this lowered the reaction capability of state regulators to the problem they were facing. State regulators could not always prevent large players from exercising market power, that is, influence the markets to their own benefit [60].

Both of these problems were at the source of the Californian Energy Crisis. In 2000-2001, the Californian Electricity System experienced very high prices that resulted in several large-scale blackouts[60]. The reaction to these event was strong, and came from all involved stakeholders. It had an impact on numbers of electricity markets and their regulation worldwide. Indeed, a massive wave of deregulation had started in many countries. No one expected that deregulation could lead to devastating consequences.

We shall not enter into further discussions about market design issues. However, we shortly discuss the consequences of the two above problems on the development of Demand Response.

In reaction to the missing money problem, some markets decided to establish a new kind of remuneration: the *capacity payments.* These consists in an insurance product similar to reserves. Generation assets that can guarantee their *capability* of providing power at peak time will get paid for it. The required total capacity is auctioned by the system operator and can take many different mechanisms. In all practical cases, the vast majority of the *capacity* is provided by generation assets. Yet, some systems have allowed demand-side resources to participate. In the US, the system in which these payments are opened to demand-side participation are CAISO, NYISO, PJM, MISO, ISO-NE [119]. This induced an enormous push for DR development. Particularly in PJM, Demand-Response faces nowadays a new golden age. The main source of revenues for demand response resources

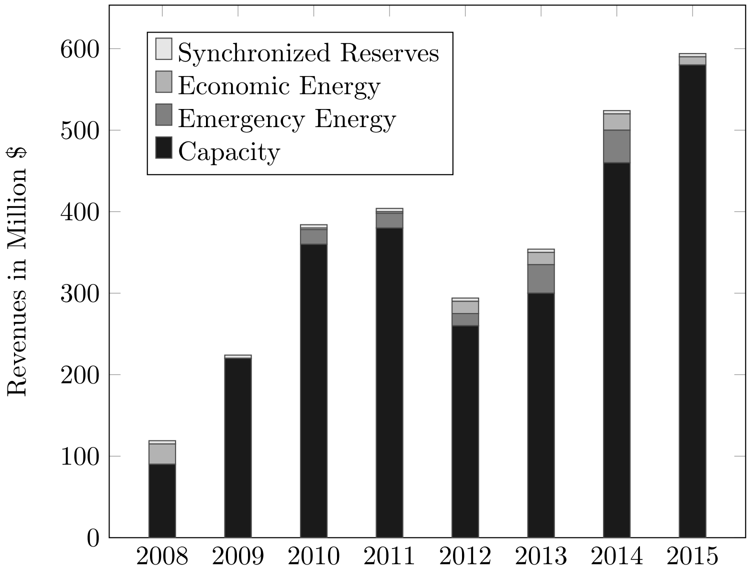


Figure 2.7: Revenues of Demand-response resources in the PJM market in 2008-2015 (adapted from [5]).

in the PJM market have been Capacity Revenues, as can be observed on figure 2.7.

Considering the market power problem, its origins are to be found in the inelastic nature of electrical demand [133]. Indeed, as demand is practically not influenced by prices, generators will get an incentive to rise the price of energy above their marginal cost. Their sold volumes will stay constant while their profits will grow. Such behavior was effectively observed in California in 2001. In response to these events, the different markets started integrating actively demand into the price formation. Economic Demand-Response programs were launched with the ambition of limiting the occurrence of price spikes and the generators market power. In response to these initiatives, the federal regulator (FERC) issued Order 745 [34] about a decade later (2011). The purpose of Order 745 was to allow demand-side resource to compete with generation offers on wholesale markets.

In order to understand the FERC order, it is important to recall that prices are the results of an optimization process. Markets solve a welfare maximization problem. Let's consider a flexible energy consumer, asking for a total of 10MWh of power. This consumer has some flexibility. He is ready to curtail 20% of its consumption in case the price rises above a 100€/MWh threshold. Let's suppose that, after optimization was conducted, the price of energy spikes at 400€/MWh. In such case, the customer will be forced to curtail its consumption as a result of the market clearing. Indeed, only 80% of its demand was cleared in the market. The rest was automatically curtailed in order to limit the price spike. The FERC order states that the curtailed portion of demand (20%) should be paid to the flexible customer at the clearing price. The customer will therefore earn 400€/MWh for not all the energy he never consumed.

Many voices raised after this Order was released. Some economist rapidly recalled that such integration of Demand would be detrimental to generators and that it was deliberately favoring Demand [124]. EPSA, the Electric Power Supply association, rapidly called the U.S Court of Appeal that eventually ruled out Order 745 [10]. Among other elements, we may read the following in [10].

> *FERC acknowledges that wholesale demand response 'selling' flexibility is a bit of a fiction.*

In Texas, the efforts to counter-strike the market momentum showed a clear trend towards promoting Demand Response. One example, more than 7 millions of smart meters are *operating* in 2014[10]. They were not even accounted in 2001.

In the case of Belgium, Demand is nowadays increasingly perceived as a reliable source of short-term flexibility as well as a capacity resource[11]. Indeed, a recent law[12] finally translates some of the elements proposed 14 years before. As example, in 2013, joined efforts of the TSO and DSOs led to the creation of a dedicated reserve product: R3 dynamic profile (Replacement Reserve). The goal was to capture the flexibility of all demand sites connected at DSO level.

2.2.5 Demand Response potential : the 10-5-5 order of magnitude.

Altogether, about 10% of a system load can be actively managed several days per year for adequacy reasons (peak load reduction). According to [77],

[10]IEA Annual Report:`http://www.eia.gov/electricity/data/eia861/index.html`
[11]`http://www.elia.be/en/grid-data/Strategic-Reserve`
[12]26 mars 2014 - Loi modifiant la loi du 29 avril 1999 relative à l'organisation du marché de l'électricité.

about 3% and 6% of the peak load in the EU and the US[13] respectively can be *shaved* using DSM measures.

On a daily basis, demand flexibility is more difficult to harness. The amount of energy (not power) that can be controllable stays below 5%. Indeed, constant day-to-day response is possible only under strong economic incentives (rare), automation (limited applications) and restricted to user's preferences. Furthermore, constant adaptation of end-use customers usually lead to *response fatigue*: users get progressively annoyed by the amount of efforts they have to provide.

Minute-to-minute response requires adapted technology. Short-term flexibility (e.g. operating reserve) can be provided by demand to a relatively large extent, in a range below 5% of the available demand [153]. This potential varies hourly and/our seasonally which decreases the economic opportunity. An exception to this last comment occurs in systems where seasonal shortages of traditional flexibility match the flexible demand availability profile. Yet, DR is strongly asymmetric. It is in general easier to decrease load than to increase it, except in some particular cases (e.g., aluminum processes delivering high value service in the US midwest [39] or Germany [92]).

Finally, when limited to absolutely critical situations (e.g., 1 in 10 years), some systems are able to disconnect up to 60% of their load in a (more or less) selective way [142], part of it at very high cost. This may only be used as a last resort resource for preventing cascading outages and blackouts.

Therefore, traditional DSM programs contribute significantly to overall system efficiency [71]. However, it is hardly comparable to the efforts that would be required for creating a *paradigm shift* : demand completely adjusting to available supply [150]. This is clearly visible in Fig. 2.5. Yet, involving energy consumers in power system operations has always been considered as very important [124]. This importance will be emphasized in the future.

2.3 Demand response in Belgium today and in a possible future.

Nowadays, Belgium faces critical adequacy issues. To counteract this trend, the Belgian state recently came up with a capacity mechanism called *strate-*

[13]The most recent assessment in the US [93] gives a potential of 12%, a figure that has be shown to be an overestimate as shown in [29] and reference therein.

gic reserve (SR). This takes the form of a capacity option. Generation assets (SGR) and demand-side (SDR) assets must guarantee that they can provide a certain capacity in case of need. Next year, a total capacity of 3500 MW (> 20% peak load) is needed (×3 last year requirement). This mechanism represents a tremendous opportunity for demand-side assets (including back-up generation) to get a steady stream of revenues.

As may be seen on Figure 2.8, Belgium's electricity consumption is dominated by the industrial sector (47%). Most of it is consumed by the chemical sector (pharmaceutical, petroleum, plastic and chemical products) closely followed by the metallurgic sector (steel, non-ferrous metal). The tertiary and residential sectors consume 26% and 24% of the total demand respectively. Shops and administrations represents most part of the tertiary sector. Breakdown shows that heating takes the largest shares of residential consumption (half of it is water heating).

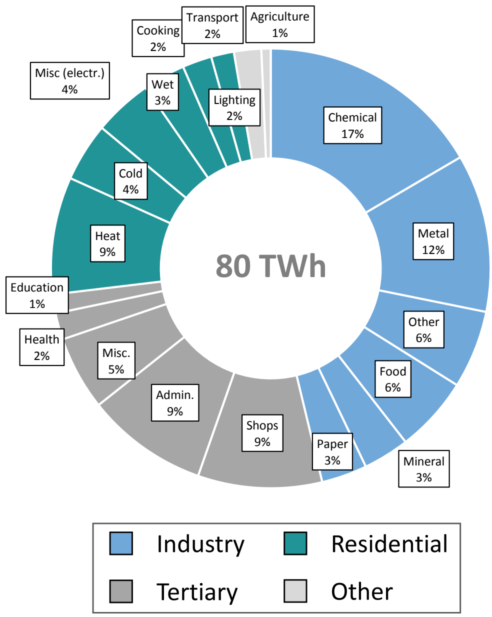


Figure 2.8: Electricity demand[14] by sector.

A recent study [63] has described the demand response potential across EU countries. The results must be considered with care. They are most of all useful to compare the different countries situation, and to highlight flexibility's typical behavior across different sectors. We will try to extract the most relevant results applicable to Belgium in this section.

2.3.1 Situation in the industrial sector.

The share of the industrial sector in Belgian electricity demand is relatively high (top 5 in Europe). Belgium has an electricity intensive industrial sector and low shares of electric heating in the other sectors. This industrial demand is exploited at the system operator level for adequacy and balancing purpose. Additionally, energy suppliers propose flexibility programs to

[14] 2014, sources: SPF, Elia, ICEDD, VITO, IBGE, [59]

benefit from price differences on the day-ahead market (planning phase) or on the real-time market (reactive balancing). At the system level, industrial loads provide a flexible capacity of close to 3.7% of the peak load (in total : 480 MW). Around 100 MW provide strategic reserves (peak load decrease), 310 MW provide tertiary control (slow reserve), and around 27 MW provide fast reserve (rare activation).

Several optimization options are proposed by aggregators and balancing responsible parties. Precise estimates of the delivered volumes are difficult to establish (confidentiality). Consumption optimization may help a BRP balancing its portfolio or be virtually sold to other BRPs in the real-time market. Indeed, the settlement scheme implemented at the Belgian level allows BRP to react to price difference and balance the system on a decentralized basis.

A bottom-up study [48] that was conducted with industries that together use up to 13.5% of the Belgian annual electricity consumption, argues that around 164 MW could be added to present situation. The total potential would reach 700 MW (of which decentralized generation is part of). A top down approach, using general figures and country comparisons in [63], finds a final potential (demand only) of 614 MW, in line with the previous result. Moreover, this latter study also gives a potential of 132 MW load increase. Load increase represents consumption anticipation. The future consumption is decreased what should be taken into account in the planning.

Generally speaking, demand response in the industry is asymmetric. Postponing/reducing consumption is easier to perform than consumption increase. In fact, industrial processes are more often in use than stopped for economic reasons. Industries for which the total costs are strongly dependent on the electricity consumption have an incentive to modulate their process for flexibility reason (metallurgy, some chemical processes, paper, cement). As the industrial electrical consumption is somehow stable in time, so is the industrial demand response.

2.3.2 The tertiary sector

The Demand Response potential in the tertiary sector mainly comes from controlling cooling and ventilation devices. Chillers in shops and administration could be controlled for delivering flexibility. Fresh-water distribution and water treatment are a second source of flexibility. Some flexibility may exist in back-up generation (e.g., hospital) or uninterruptible power supply in the IT sector (e.g. servers). In extreme cases, lighting are a last source of flexibility. The flexibility of cold storage warehouses and water pumps is already partly exploited today by few aggregation companies. There is no official estimation on the activation profile of these resources today in Belgium.

Excluding the potential of back-up generation, the theoretical total peak load decrease potential of the commercial sector could reach -655 MW (±5% of Belgian peak)[63], while the simultaneous load increase could reach +580 MW. Interestingly, the tertiary sector's potential is almost symmetrical, at least in case flexibility is called for short duration.

This potential is most probably unreachable in practice. Firstly, it represents a direct sum of potential from different sub-sectors (e.g. ventilation in buildings, cooling in retail,etc.). Simultaneous call of the whole tertiary sector's flexibility would in practice deliver a lower level of power than separate calls from its different sub-sectors. Secondly, it is evaluated at the most favorable time. Finally, it is theoretical. All loads are therefore considered externally controllable and economic considerations as well as control acceptance by users are neglected.

2.3.3 The residential sector

Traditionally, DR programs in the residential sector have been designed to control single load types (AC, water heaters), delivering high-value services (frequency support, peak load decrease) and used in rare events. Residential consumption is composed by a very large number of small appliances, active for short duration. Therefore, it is extremely complex to apply tight control on their consumption. The relative use of a specific appliance is low which leads to a strong asymmetric flexibility potential ([63] indicates a total potential of -782 MW/+4023 MW). Indeed, it is (technically) much easier to anticipate future energy consumption than postponing current consumption, certainly for very short time. In practice, when individual load consumption is planned by the user with some flexibility, it becomes possible to shape the residential load profile.

The Linear project recently explored residential consumption flexibility in Flanders. While this project is not the first pilot project of its kind, it has still elegantly highlighted typical characteristics of residential demand-response. In [97] (pp.85), it is shown to be technically possible to decrease the residential consumption of 267 MW during 4 hours without impacting the users' comfort. This potential mainly comes from the so-called *smart water heaters*.

The main problem of residential demand control is its low per-appliance profitability. Firstly, when grouped by appliance type, the consumption profile very often varies across hours and seasons [79]. Only fridges and freezers are used on a more or less constant basis. The value of the residential flexibility is very often negligible compared to the initial energy purchase costs. Thermal loads are typically able to save less than 5% of their initial costs.

In the long-run, electric vehicles could change the picture. Their energy needs are extremely high, and their charging can be controlled to be less damageable to the system. Also here, simple programs should be preferred to prevent response fatigue (see below). Furthermore, it has been proven that results of simple programs can be very close to complex ones [128].

2.4 The road to success for demand response programs.

It is possible to fail in many ways, while to succeed is possible only in one way.

Aristotle - Ethika Nikomacheia (350 BC)

DSM is a required and beneficial feature in restructured energy markets. Some very good example exists, showing that demand can contribute to system reliability, price stability, risk mitigation, and market efficiency. However, some other examples have shown the limit of DSM.

Essentially, the long term success of a DSM program can be guaranteed under two ways. Either the program is cost-effective and it will bring more benefits (or avoided costs) than what it costs, or the program is *subsidized* by a specific entity (market player, government, etc.) [124]. In the first case, the benefits that will be achieved will maintain a continuous attractiveness for the DSM program, such that the long term success is insured. In the second case (subsidized program), a strong market player (i.e. vertically integrated utility) will have to tax consumers in order to insure profitability of the DSM measure. The program will exist only if this market force is maintained.

When promoting a DSM program, a crucial attention should be taken by decision makers to, either, the cost effectiveness, or to the long-run market/system structure in which the DSM measure is implemented (i.e. there should be sufficient control on the

> *market/system such that the subsidy is maintained). This will ensure, in two different ways, a long term success to the program.*

According to [130], effective DSM programs are expensive and labor intensive. In today's situation, part of the envisioned demand response potential still does not pay off, particularly in the residential sector. Demand response will stay deployed in systems where its use is vital for system security and stability, with restricted participation to infrequent events, and where the power system demand has significant shares of temperature-sensitive loads or large industrial processes. Price differences are today not sufficient to foster demand participation. This will definitely change in the near future. However, a longer term outlook indicates that increased price volatility will make it hard for consumers to follow price signals. There are several answers to the problem.

The most commonly shared answer is technology. Smart-metering and control infrastructure will help delivering their full potential while the end-user will not be directly involved. However, several studies have shown that the comfort-constrained potential of demand-flexibility is somehow limited [88, 90, 103, 22]. In some years, together with the development of fossil-fuel alternatives in the transportation sector, energy storage could become affordable at small scale level (hydrogen/fuel cells or electric batteries are promising, even though material/water availability can change the picture). DSM-like programs could be adapted to take control of the decentralized storage capacity. Furthermore, as we will show below, long term involvement is crucial for the viability of demand response programs. It is therefore worthwhile starting now.

But technology alone will not make it[15]. Improper smart-meter's information has a negligible impact on users' consumption, and may even increase consumption. Success is associated to behavioral sciences, market and policy design rather than to pure technical considerations. Transparency and simplicity are key to favor massive participation.

Truthfully, it is hardly possible to generalize one system situation to an other. Indeed, technical, economic and most importantly social and cultural factors influence the success of a certain program. Knowing this, if some key conditions are not present, the program will fail.

[15]Interesting article of the Washington Post about the US situation: `http://www.washingtonpost.com/news/energy-environment/wp/2015/01/29/americans-are-this-close-to-finally-understanding-their-electricity-bills/`

2.4.1 The requirement of long-term commitment from all stakeholders.

Demand response programs are very sensitive to change. A changing *market* environment discourages demand to actively participate in power system operations. The liberalization process that has recently taken place in numerous systems worldwide is a good example. The responsibility shift lead to decreasing DR interest in the eyes of the newly created players (e.g., Texas: [157]). In PJM, changing rules in the remuneration scheme and pre-qualification tests resulted in a fade out of the demand participation rate in wholesale markets (economic demand response). The low gas prices in the US may also have played a role, due to lower price spreads [16].

In Belgium, the liberalization engendered the creation of a TSO and several DSOs. Some DSOs became responsible for portion of the network to which demand sites previously exploited for balancing purposes. Due to the changing environment and no vital reliance of the system to these demand resources, these small sites became unexploited. Only recently, the involved stakeholders have agreed on rules to integrate DSO-connected demand into TSO's operations (balancing). Such lessons are very important if Belgium had to rely on DR to guarantee safe operations. The DR program will need to stay unchanged in the long term.

2.4.2 Following a precisely defined objective.

Successful DR programs always follow a very precise objective : peak load decrease, economic opportunity, etc. As discussed above, the underlying motivations to develop DR may come from all interested players : network operator, energy supplier, market operator, energy consumers.

2.4.3 Well-balanced design

Well designed demand response programs fulfill several conditions for success. A balanced DR Program should (1) address the suited part of flexible demand taking into account demand-side costs (2) use the simplest imaginable program (3) ensure profitability without subsidies meaning that total system costs should decrease with the introduction of the program (4) avoid conflicts of interest for resource access between the different players (5) adapt rules and technical requirements to cope with demand intrinsic limitations and (6) keep motivating the *demand resources* actively.

[16] Source: `http://www.energymanagertoday.com/`

Address the suited loads and the appropriate consumers.

Flexible loads are usually able to perform different kinds of services. However, some services are more suited to their technical or economic characteristic. Flexibility costs are relative to consumer's preferences (habits), to the inherent technical nature of the electric load and to the added-value that its use may bring.

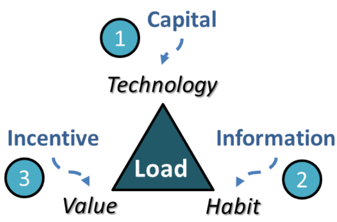


Figure 2.9: Load's threefold nature.

As pictured in Fig.2.9, electric appliances/processes have a threefold nature (1) a technology (2) a value and (3) a habit. Any external intervention on demand impacts one or more of these natural features inducing direct or indirect costs as well as opportunities. Technology is influenced by capital cost (investment in new appliance), habits may change with information and value will depend on external incentives. Consumers evaluate and compare all three to decide on their level of participation to a certain program.

Addressing the right technology is important. Demand flexibility has four technical origins: (1) redundancy (e.g. back-up generation, process security) (2) pure storage capability (e.g. water distribution, night storage heating) (3) comfort/quality-constrained storage capability (e.g. fridges) (4) forgoing consumption[17]. These technical origins lead to different kinds of flexibility (energy, power and dynamics).

The economic value extracted from a load is sector dependent. Industries have stable consumption and asymmetric flexibility (high use of installed capacity). Residential consumption is more variable and strongly asymmetric (low use of installed capacity). Tertiary sector is somehow in between of these extremes.

Load with similar technical characteristics and from which users extract a comparable value can still be used differently according to a user's habits. These may be influenced by information. However, this information must be carefully chosen and displayed in various forms, as explained below.

Addressing the right consumer is also important. Badly designed programs face an adverse selection problem : recruited consumers are those that

[17]Forgoing loads may be occasionally switched off for short duration without significant impact on the user's process but have no storage capacity (e.g. light)

have the lowest capability to react. This phenomenon is often observed in price-based program implemented to solve adequacy issues. Consumer with low peak-time consumption have higher incentive to participate (i.e., they face less risks and benefit more) while they are the less suited to react. Therefore, the *pricing vs control* dilemma basically defines to which degree price risks should be passed on to consumers [58].

Simple programs for long-term success.

The success of a DR program is measured at the level of sustained consumer involvement in a long run perspective. It has been shown that complex programs are likely to discourage users from participating. There is evidence that users progressively lose their involvement in programs that they do not understand perfectly [77]. This effect is called *response fatigue* and is very damaging in practice, as it can progressively wipe out the benefits of huge efforts made to access demand flexibility.

In general, the provided information should allow for progressive learning. Indeed, consumers like to *explore* and learn from past experience. Information should be available in various forms and allow for social comparisons. Users and habits are strongly influence by social interactions. Price is not the most important trigger to attract massive consumer response [96].

Profitability and costs allocation without subsidies.

Demand response relying on direct or indirect subsidies are not sustainable. As well explained by Paul Joskow in [75], it is not worth subsidizing efforts that consumers would have done under proper pricing programs or with adapted information. An effective program relies on users' self-interest which can be fostered by both information and incentive.

The problem of subsidies is not necessarily the immediate financial distortion it creates, but rather the influence it may have on the long-term system equilibrium. In some sense, in today's system, users who contribute the most to peak demand are *cross-subsidized* by the ones having a flatter demand. Dynamic pricing essential goal is to diminish cross-subsidies and allocate costs more precisely. Yet, it should be used with care as it may result in large payments if consumers are unable to react to price signals.

Sharing DR between multiple players : avoid conflicts and protect data.

Demand flexibility is a resource that may be accessed by very different players. Furthermore, the multiple intermediaries' involved in power system operations may all be impacted by demand response actions. It is important to organize the market design, payment schemes and compensations such that the foreseen demand actions do not enter in conflict with the different intermediaries interest. On the other hand, an intermediary should not profit from actions undertaken by an other. There is however a complex dilemma. Indeed, data transparency is key for verification (cross-checking) purpose while at the same time it can rise privacy issues on the consumer side, and strategic concern on the flexibility provider's side (w.r.t its competitors).

Adapt technical requirements.

Demand resources do not have the same capability as generation assets. Small loads can be imprecisely controlled while large processes often face energy limits. A program should consider those constraints and adapt the technical rules to alleviate their impact in the most efficient way. Rules should neither prevent demand to participate in services it can extract value from, nor should it be too favorable to demand assets.

Technical requirements also encompass standards, and forms of mandatory participation of demand. Some loads (e.g., see next chapter) could be used to provide very fast and valuable flexibility. Some pilot programs have shown very cost-effective. Mandatory participation and standardization should also be considered with great attention.

2.5 Chapter Conclusion.

Nothing is so painful to the human mind as a great and sudden change.

Mary W. Shelley - Frankenstein or the modern Prometheus (1818)

In this chapter, we showed that Demand Response has been able to deliver a broad panel of services. The access to Demand Side resources is complex by nature, as it requires the agreement of many different parties. It is therefore mostly developed in systems where it provides a service of high value.

In Belgium, and today's situation, deploying demand response in a cost-effective way can provide half of the required short-term flexibility (reserve) and diminish peak load consistently.

In the long run, the envisioned shift to solar-based energy will completely change the traditional organization/operations of power systems. The change is so profound that it is hardly possible to imagine, based on current practices. Large infrastructural changes are needed both on the network-side and on increased decentralized storage capability (thermal storage, batteries, more flexible industrial processes) before demand-side response can effectively alleviate the variability of solar-based generation.

In Europe, the market will most probably evolve with the increasing needs for storage capacity. If DR had to cover a large portion of those needs, the *long-term* commitment from all involved stakeholders, the design of well-balanced demand response programs and the definition of clear program objectives will be crucial for realizing its full potential.

Chapter 3

Frequency Containment Reserve from Small Electric Loads

Chapter summary

One day a terrible fire broke out in a forest. Frightened, all the animals fled their homes until they reached the edge of a stream, watching the fire, helplessly. Every one of them thought there was nothing they could do about the fire, except for one ***small*** *hummingbird. It swooped into the stream, picked up a few drops of water and went into the forest to throw them on the fire. Then it went back to the stream and did it again, again and again. All the other animals watched in disbelief. Then one of them challenged the hummingbird in a mocking voice, "What do you think you are doing?". And the hummingbird answered: "I am doing what I can".*

Adapted from Wangari Maathai, NAFSA 58th annual conference, May 2006[a].

[a] http://www.wangfoundation.net/humming_bird.pdf

This chapter explores the historical efforts that have been spent worldwide for exploiting the flexibility of small electric loads within system balancing and ancillary services. On the academic perspective, recent advances in computational tools have made large scale simulations possible that adequately represent their behavior and assess their potential.

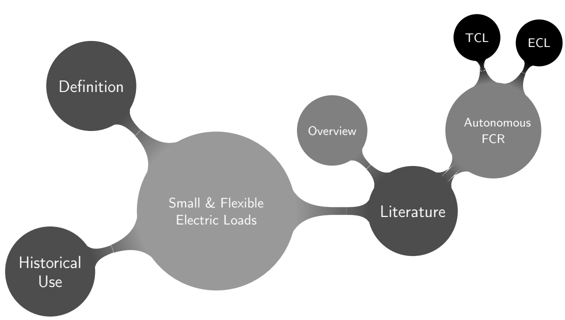


Figure 3.1: Another view of this chapter

3.1 Small electric loads: a definition

Small electric loads gather all electric appliances whose power rating is below or equal to 25kW. As example, let's imagine all pumps, compressors and fans, motors, heating coils that are used in heating, cooling and compression processes. Also, electric batteries used for energy storage purpose, redundancy or power quality issues are part of this class. The energy consumption of small loads represents important shares of the total EU-wide electrical energy consumption. Furthermore, they are generally associated with physical processes with intrinsic energy storage capability. Would it be because of security reason or due to the end-user behavior, these loads are often under-utilized. That is, the proportion of time in a year during which they actually consume energy is rather small. Hence, their user or the associated process allows in theory to use those loads at a different time than the one initially scheduled. Altogether, their consumption can be externally controlled. The main challenge consists in finding ways to efficiently transfer information between the end-user and anyone that would be interested in influencing its consumption.

3.1.1 Small loads in the EU-wide electrical energy consumption

In 2013, about 58% of the EU-wide electrical energy consumption consisted in running simple motors, pumps or heating elements for generating cold or warm air and water, transporting intermediary products within industrial processes or storing compressed air. Certainly half of this consumption can be attributed to loads with a nominal power below 25 kW. Indeed, most of the loads installed in buildings and used for space/water heating or refrigeration enter this category. As illustration, different pie-charts are presented below. The colored shares represent the portion of demand in which small loads are likely to represent most of the energy consumption. Their exact shares is hardly estimable, though, especially in the industrial sector.

In the EU-28 area, the final consumption of energy reached 12.8 PWh in 2013, of which 2.8 PWh (22%) were in the form of electricity. The electricity consumption was distributed as follows among the different sectors : 36% Industrial, 30% Service and 30% Residential, and 4% Other use. The EU electricity consumption breakdown by end-use is represented below for each sector: residential (Fig.3.2), service (Fig.3.3) and industrial (Fig.3.4). The darker shares in the Residential sectors are populated exclusively by small loads. So is it for the commercial sector, except part of pumps (e.g., large pumps in drinking water production) and refrigeration shares (e.g., large coolers in warehouses). The cooking and lighting shares are excluded one may hardly interfere with their use. In the industrial sector, more detailed analysis should be conducted on each industrial branch to separate larger loads from small loads.

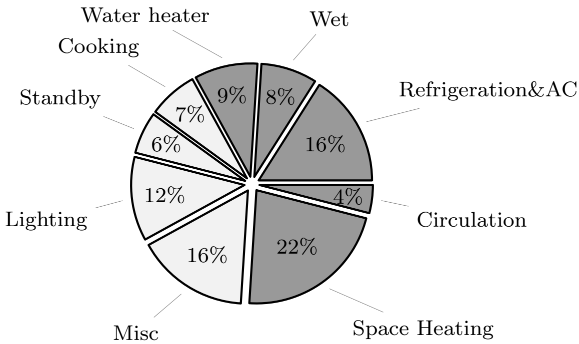


Figure 3.2: EU Residential Electricity consumption breakdown (2010), [11].

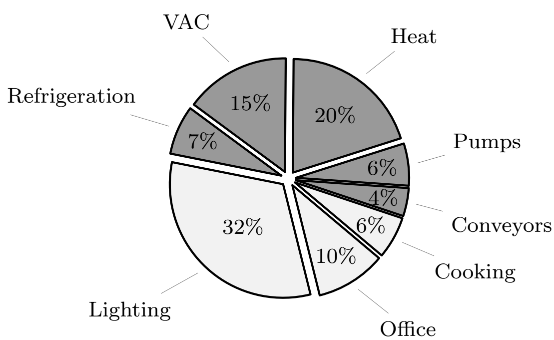


Figure 3.3: EU Tertiary Electricity consumption breakdown (2010), source: [11].

Several studies have looked at potential of using small loads in system operations. A lot them have been conducted on the German System [131, 80, 78, 118]. According to [41], controlling 60% of the cold and ventilation appliances in the industrial and commercial sector of Germany would be sufficient to cover all the replacement reserves at a cost between 100 to 400 €/MW (partially competitive). According to [69], the loads that are the most suited to provide operating reserves are the following : Batch-type (intermediary storage) industrial processes, Cold Warehouses, Electric Water Heaters, Dual-fuel Boilers, Buildings with sufficiently large thermal

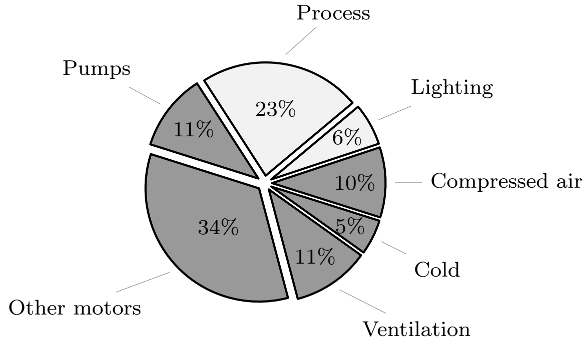


Figure 3.4: EU Industrial Electricity consumption breakdown (2013), based on: `http://www.odyssee-mure.eu` (consumption by branch) and `http://www.leonardo-energy.org` (breakdown by use).

mass. We may have to add electric vehicle and battery storage, depending on the future evolution of these markets.

3.1.2 Recent interest for integrating loads within system operations

As from 2012, the ambition of integrating small electric loads within system operations in Europe as been precisely expressed in the ENTSOe demand connection code [51]. More specifically, the *Temperature Controlled Devices*, known in the scientific literature as Thermostatically Controlled Loads, should participate *mandatorily* to frequency control. Yet, a condition is associated with this obligation: their contribution should be *significant*. In other words, demand-side control should be *efficient* in a socio-economic manner.

> Demand Side Response System Frequency Control (DSR SFC) shall mandatorily apply to new Temperature Controlled Devices identified as significant [...].
>
> Article 21, [51]

This is still very vague, and has most probably be formulated so on purpose.

Indeed, harvesting profitable response from small loads is a challenge. Firstly, accessing load's flexibility is expensive. Indeed, small loads are often connected at the far end of the network and their environment is highly constrained by user preferences. Secondly, the per-load expected benefits are limited. Programs exploiting small loads are restricted to use low-cost control and communication infrastructure.

These renewed interests in demand-side response find their origin in the operational challenges that are ahead: decarbonated power systems are likely to face stability issues. In the years to come, numerous large generation assets will get decommissioned[1]. This lead to two negative impacts. Firstly, the total rotational inertia of the system will decrease. Rotational inertia is a central feature in maintaining frequency stability (see previous chapter for considerations on frequency). This inertia comes from the electro-mechanical coupling between large generators' kinetic energy and the system frequency [147]. The higher this inertia, the more robust is the system. Secondly, large generators are equipped with speed controllers that are used to provide damping to frequency excursions around nominal value.

In a decarbonated system, the initial signal creation (nominal frequency) will still be performed by rotating masses (e.g. wind turbines). Increased reliance on load control would insure short-term system stability [132] and limit the use of storage assets for this specific purpose (e.g., batteries [86], flywheels [99]). Indeed, large rotating loads (e.g., motors) also participate in to the overall system inertia. However, these loads are increasingly fed by power-electronic devices for power quality or control purpose. Thereby, the electro-mechanical coupling is lost. Their kinetic energy is not accessible to the system. On the other hand, inductive loads provide damping to frequency changes [6].

Two main types of actions can be performed on demand. First, preventive actions (ahead of delivery time) can impact demand for relatively long periods (up to several hours). Then, reactive actions can instantaneously modify the real-time electric demand in order to balance the system on a second-to-second basis. These actions are potentially competing with each other: displaced loads may not contribute to system stability at the time they were supposed to run.

3.1.3 Historical efforts for accessing small loads.

Historically, only a few systems have been able to exploit efficiently small loads in their operations. These are often systems in which the exploited loads are the main source of the problem their flexibility helps solving. As examples, that we further discuss below, let's mention residential air-

[1] E.g., nuclear decommissioning https://ec.europa.eu/energy/en/topics/nuclear-energy/decommissioning-nuclear-facilities

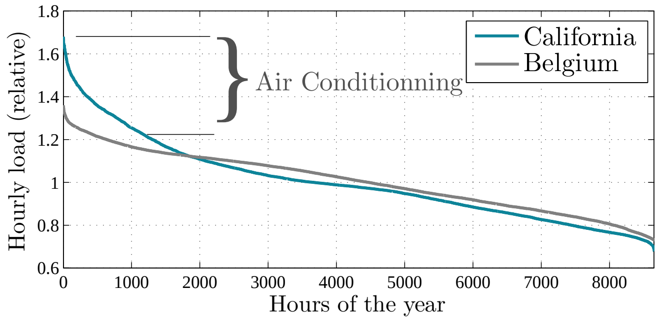


Figure 3.5: Load Duration Curves (2014) of California (CAISO) and Belgium (Elia). Hourly load values are displayed in proportion of their annual average. In CAISO, 30% (=0.5 on the chart) of the peak is attributed to air-conditionning. Source : CAISO, Elia.

conditioning (AC) in some US systems or electric heating in France. The aggregated consumption of AC or electric heating rises up seriously in severe climate condition. In those systems, the annual peak demand is therefore relatively high compared to the average annual load. This induces an expensive adequacy problem which is best solving by targeting its root causes: the loads themselves.

As illustration, the load duration curves of 2014 in both California[2] and Belgium[3] are shown on Fig.3.5. The hourly load is expressed in both cases relatively to its annual average. Compared to Belgium, the Californian system has a much higher relative peak. About 30% of that relative peak (=0.5 on the chart) can be attributed to small air-conditioning units. This explains the historical development of AC-related programs in the Californian system [19]. Obviously, this cannot be translated to Belgium.

Other US systems are subject to high peak-to-average ratio, particularly in recent years. As EIA explains[4], this is due to an increased use of electricity for heating/cooling purpose (climate control), a change in the pattern of electricity use (e.g., energy efficiency), and the shift from industrially-based economy to service-oriented economy.

[2]Data extracted from : `http://www.energyonline.com/Data/GenericData.aspx?DataId=18`

[3]Data extracted from : `http://www.elia.be/en/grid-data/data-download`

[4]`http://www.eia.gov/todayinenergy/detail.cfm?id=15051#tabs_SpotPriceSlider-8`

For these and several other reasons (unusual system topology, climate-sensitive generation mix) security and adequacy issues gave the incentive to exploit the flexibility of small electric loads. The developed programs have exploited tools ranging from price-based programs [46], to direct load control [50]. In Europe, one of the most successful program has been the Tempo Tariff launched by EDF in France [141] (and detailed previously).

In Detroit (DTE), the control of water heaters was started as early as 1934 [68]. Time clocks were used to trigger temporary disconnection of the controlled loads. They however got problems with defining the trigger time because of changing behavior at peak time. In 1968, the switching was therefore *being done via radio signal under the control of the system operator* [68]. This provided 200 MW in the winter time and still 50 MW in summer. Interestingly, the study looks at the impact of external action on the load behavior. Energy rebound was deeply analyzed and integrated in the radio signal.

In the late 1970s, another source scrutinized technical improvements of air-conditioning for increased storage capability. Authors in [15] discuss use of phase change materials for cold storage in off-peak period. Their interest originates from the previous decade in which massive investments where directed to thermal storage in both US and Europe. In the case of Europe, heating loads were inducing large demand peaks in the winter, but also gave some opportunities.

> *Both Britain and West Germany have approximately 150,000 MWh of off-peak heating storage. In Germany, this storage represents a 40% reduction in system peak load. [..] Britain has a program combining tariffs for cheaper off-peak electricity and the development of special electric products tailored to customer needs.*[15].

The authors continue by specifying that a 95% customer satisfaction rate has been achieved thanks to political willingness, government support and advertising the role of these new appliances.

Following the energy crisis of the 1980s, Florida Power&Light Co. developed a program in which small electric loads (e.g., AC - Air conditioning , water heater, pool pumps) were automatically shut down 3 to 4 times a year [8]. Up to 816000 appliances and 712000 users got involved to deliver 1000 MW of flexibility. During the early development, about 1.5% of the installed controllers generated calls from customer. Some were worried that the controlled appliances would experience a larger rate of failures. The program paid each month 6$ for AC and 3.5$ for water heaters. These relatively high payments had a single purpose : insure generation adequacy at peak and shoulder hours (maintenance).

In [105], authors mention the use of demand limiters in 1985 applied on thermal loads. Commercially available power limiters were modified to be remotely controllable (different signals would limit demand to 33%, 55%, and 1 for 15 minute interval).

> *The observed effects were highly weather-sensitive, largely because the controllable load is driven by the ambient temperature and other weather-related variables. The demand limiters were more effective as the outdoor temperature approached extremes. However, they became less effective in the later part of the winter seasons. Frequent direct control by the utility to reduce demand was found to negate the effect of the local logic.*

3.1.4 Revamping the concept of Energy efficiency ?

Flexible small loads may soon become an important contributor to the electricity system's short-term stability. Appliances whose consumption is controllable and whose energy needs can be deferred in time have two main characteristics: on-board intelligence (for control) and storage capacity. Their design is a crucial element that will condition the amount of flexibility that can be delivered.

Our traditional way of thinking about energy efficiency may lead to designing appliances consuming very little energy, but with low flexibility. Policy makers will soon have to struggle with the following question. What appliance is the most energy efficient: the one that consumes the less or the most flexible one? Let's note that these are not mutually exclusive features, but this depends on the specific appliance type and usage.

For instance, increasing the insulation that surrounds a water boiler will increase the flexibility potential as well as decrease the average heat losses. However, increasing the water temperature within the boiler's tank for flexibility purpose, for instance in case the energy consumption has to be anticipated, will induce slightly higher losses. In addition, boiler oversizing leads to higher material/energy use during the manufacturing process. Grasping the exact consequences of these design choices is therefore very complex.

3.2 Small flexible loads and frequency containment reserve: literature.

The literature is, literally, overflowing with studies that consider small loads as a valuable source of short term flexibility. Most of recent work focuses on simple ways to interact with large loads groups. Individual load's consumption can be modeled with growing complexity from a simple energy constraint to a detailed dynamical system. The scientific challenge consists in understanding the behavior of the group when its elements are individually controlled. Research areas cover modeling, control/optimization algorithms, cost-based analyses and (rarely) pilot projects assessment.

The literature explored below addresses specifically the use of small flexible loads within frequency containment reserve (primary frequency control). Some papers have been approaching the subject across all dimensions (see e.g., [72], [28]). The following review is not designed to be exhaustive, but rather very precise. We want to illustrate efficiently the state of today's knowledge on the integration of small electric loads within Frequency Containment Reserve.

The selected elements of the literature are firstly classified according to (1) the type of flexible load studied, (2) the control setup and (3) whether the group dynamics (short-term time evolution) are taken into account. We give an overview of interesting elements of the literature on Table 3.1. The most interesting ones are discussed in details in the next section, while the classification used in this table is explained below.

Table 3.1: Literature - Small loads providing frequency control.

Control Setup		Centralized	Decentralized	Autonomous	
Group-level Dynamics		With	With	With	Without
TCL	Individual	/	[140, 126]	[159]	[127, 111, 44, 37] [156, 9, 18, 146]
	Mixed-type	[62, 3]	[158, 81]		
	Aggregate	/	/	[24]	/
ECL	Individual	[17, 85, 16] [86, 102, 152]	[84, 112]	/	[98, 144, 45]
	Mixed-type	[13, 62, 45]	[108]		
	Aggregate	[68]	/	/[a]	/[a]

[a] The object of this work.

As can be observed in the above table, the literature did not cover all possible combinations of the sorting criteria. Firstly, let's insist, this table is relative to studies looking specifically at frequency control. Secondly, some of the combinations make little sense, and are darkened in Table 3.1. Indeed, mixed-type models are used for computing control commands and become unnecessary when loads are controlled autonomously. Thirdly, some gaps are partially filled in the literature looking at price-based dispatching of small loads. Finally, some other have not been explored and are the object of this work.

Individual load types and models

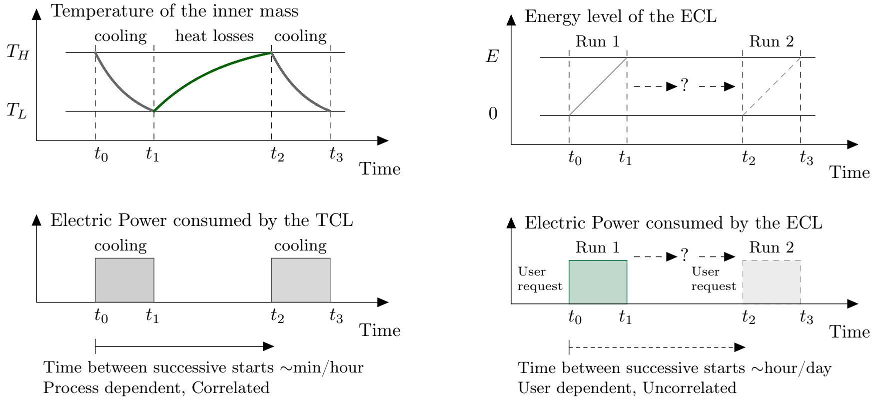
When and how often does a load start ? And when started, how does its power consumption evolves with time ? Does the load stop ? And when it does, when and why does it start again for another run ? More importantly, how can its power consumption be controlled and what are the consequences ? These are the typical questions addressed by individual load models.

The flexible loads that are studied in the literature may be classified into general load types or specific load types. General load type models represent loads with some typical behavior (e.g., cyclic consumption profile), while specific load type models explore some special kind of loads in more details (e.g., heat-pumps with specific technical constraints). Two general load models exist in the literature: thermostatically controlled loads (TCLs) and energy constrained loads[5] (ECLs).

The TCL model is a general representation of loads whose power consumption aims at regulating the temperature of an inner mass. The model of TCL encompasses cooling appliances such as refrigerators, freezers or air-conditioning as well as heating appliances like some heat pumps. As illustrated on Fig.3.6a, the power consumption of a TCL follows a cycling on and off pattern that is governed by heat exchange equations and some hysteresis behavior. Indeed, a thermostat is used to trigger a load start as soon as the measured temperature goes beyond a first threshold (e.g., T_H higher temperature limit for cooling appliances). The load is later switched off as the regulated temperature pass a second threshold (e.g., T_L lower temperature limit for cooling appliances).

The ECL model gathers all storage-like appliances: batch water heater (e.g., sanitary needs, dishwashers, washing machines), electric vehicle charging, fluid pumps in batch processes (e.g., pool pumps, water sanitation, beverage industry). An ECL starts consuming at constant rate until it has consumed a predefined amount of energy E. The load then stops until restarted by another user request. As shown on Fig.3.6b, successive starts

[5] The ECLs are also known as limited energy resources or energy constrained resources.



(a) Two successive runs of a cooling TCL. (b) Two successive runs of an ECL.

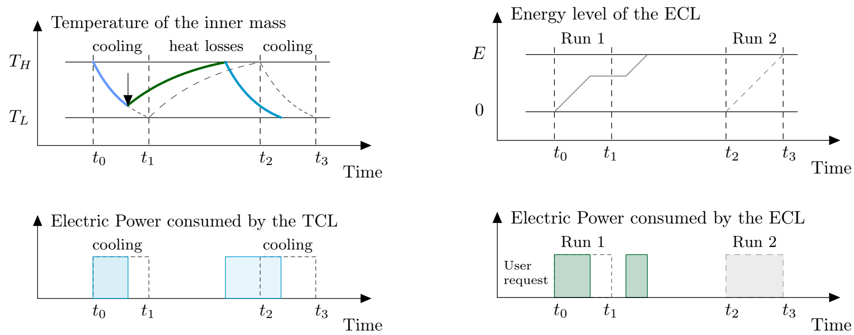
Figure 3.6: Two general load models and correlation of successive starts.

of an ECL are, statistically speaking, mutually independent.

Figure 3.6 illustrates the concept of successive *runs* of both TCL and ECL. A load *run* is defined as the period of time in which a load either consumes power continuously or is externally idled (i.e. its power is deliberately switched off). For a TCL, a run corresponds to a period when it consumes power. For an ECL, it is the period of time between the user starting request and the time at which the load has consumed the requested energy E.

Generally speaking, a small electric load consumes energy either as consequence of a direct request of its user or following a command issued from its internal state control. For instance, dishwashers (ECLs) have a user-related starting behavior while refrigerators' (TCLs) starting times are linked to the evolution of the inner mass temperature they regulate. From a static perspective, these two starting triggers are random variables. Indeed, counting down in a large group the number of loads that are running at some point in time boils down to evaluate some static probabilistic distribution. A system operator can thereby estimate, for example, the power consumption of a large group of similar loads.

However, this does not hold in case loads are externally controlled to deliver flexibility services. Indeed, the *succession* of several starts of a single load can be strongly time correlated. More rarely, they can be the consequence of repeated user behavior. Usually, such repeated pattern in user's



(a) External control on cooling TCL.

(b) External control on ECL (Run 1).

Figure 3.7: Impact of external intervention on successive cycles.

behavior span on long time periods (i.e., days), which is why they can be neglected in case loads participate to short-term flexibility. Typically, a TCL's successive starts are strongly correlated (Fig.3.6a) as their temperature evolution is driven by heat exchange equations. On the other hand, successive ECL's starting times are considered uncorrelated (Fig.3.6b). Correlations naturally emerge when differential equations influence the load's behavior.

Correlations introduce complex dynamical behavior at the group level. External interventions on the load's behavior will have immediate as well as longer term impact. The longer term impact is the consequence of starting time correlations as well as the loss of demand diversity. The first element can be observed on Fig. 3.7. On Fig.3.7a, the first cooling cycle of a TCL is externally interrupted. As a consequence, the regulated temperature starts increasing while it had not reached its lower limit. It therefore reaches its higher limit sooner than what would have been observed without the initial intervention. The following cooling run starts earlier than expected. Contrastingly, cross-run interactions do not arise in the case of ECL (Fig.3.7b).

The loss of demand diversity is a group-level feature. It cannot be apprehended from individual considerations.

Let's note that some thermal loads have an intermediary behavior. Even though their temperature is regulated, the time at which they start consuming is strongly related to external events. For instance, well-insulated storage heaters (e.g., ceramic storage, water heaters) consume most of their energy at night or low-price periods. The load generally starts heating up its inner mass and then stops when the temperature has reached a certain threshold. Successive starts can be observed during high price period. Yet, they are the consequence of a user action (e.g., water draws, etc). On a system perspective, the user behavior is the dominant driver of their consumption.

Aggregation level and aggregate model

Small flexible loads must be studied as part of large groups in order to assess the system-level consequences of their control. Hence, the above individual elements must be gathered together in some ways. This is the role of aggregate models. At this stage, and for the sake of clarity, a distinction is worth being made between the terms aggregate flexible capacity and aggregate model.

Aggregate flexible capacity refers to the capacity (in MW) provided by a group of loads when performing a specific service. All studies discussing controllable loads will scrutinized how much aggregate flexible capacity can be extracted from the group.

Aggregate models are mathematical structures that are built to represent the *dynamical* behavior of a large group of controllable electric appliances, with possibly different parameters. Aggregate models reduce computational efforts required to simulate and/or optimally control the power demand of large groups of loads [40].

Both terms are sometimes mixed-up in the literature. This is because the most simple form of aggregate model consists in supposing that the aggregate flexible capacity has similar dynamics than an ideal generator (e.g., [73]), which is a rough approximate. In the literature, studies differ from each other depending on the *aggregation level* they consider. That is, whether the study considers each load separately or directly exploits aggregated-level information.

There exist three aggregation level: individual, aggregate and aggregate with feedback coupling (mixed model). At the individual level, each load is separately modeled and simulated. The aggregate level rely solely on the aggregate model to simulate the group's response. Such models are particularly useful. They often allow to run computationally affordable simulations on long periods of time which can be of importance to e.g, economical analysis. At the interface between these two cases, some aggregate models proposed in the literature do not capture all features of the individual models. These mixed-type models designed for short-term control purpose (e.g., linearized model) and must be coupled though information feedback to individual simulations.

A few assumptions are needed to define four essential features of aggregate models: (1) the model states, (2) the model parameters, (3) the control inputs and (4) the external perturbation. The model states are time-varying variables which evolution are driven by the models equations. The parameters are exogenous quantities that appear as fixed elements in the model. The control inputs define the influence of any external intervention on the system states. External perturbations are all elements that influence the system states in a undesired or uncontrolled manner. On Table 3.2, the states, parameters, inputs and perturbations relative to both TCL and ECL models are shown.

Table 3.2: Assumptions governing the aggregate models of ECL and TCL.

Load Type	States	Fixed parameters	External perturbations
TCL	Inner Temperature Electric Power	• Temperature limits • Power profile • Number of loads	• External temperature • User action
ECL	Energy level Electric Power	• Energy need • Power Limits • Time deadline	• Number of load starts

Degree of decentralization of a control setup

The degree of decentralization of (1) the loads command and (2) the information feedback [70, 25] defines the control setup.

In a *centralized control* setup, each load is accounted and controlled individually by a central entity. The loads are equipped with a **two-way communication** system that allows them to receive and send information (Fig. 3.8a).

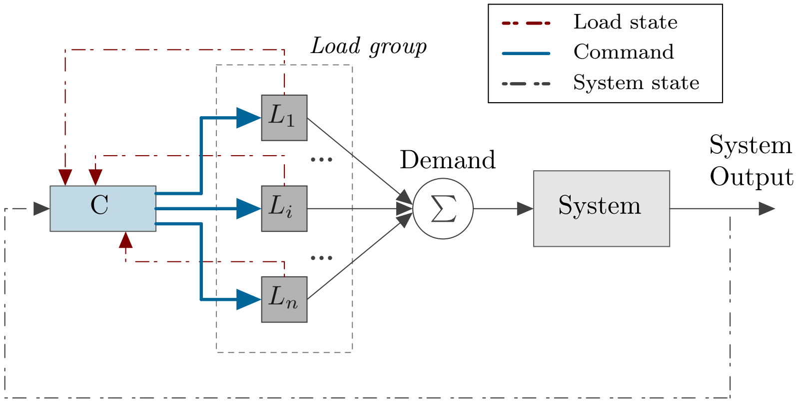
In this work, we denote as *decentralized control* a control setup in which loads are equipped with **one-way communication** system. Loads receive a communication signal from a central entity (e.g., price signal) that they locally interpret and convert into control inputs (Fig. 3.8b).

Autonomous control defines a setup in which loads take control decisions individually using a local controller that rely exclusively on locally measured information (Fig. 3.8c). No extra communication layer is considered.

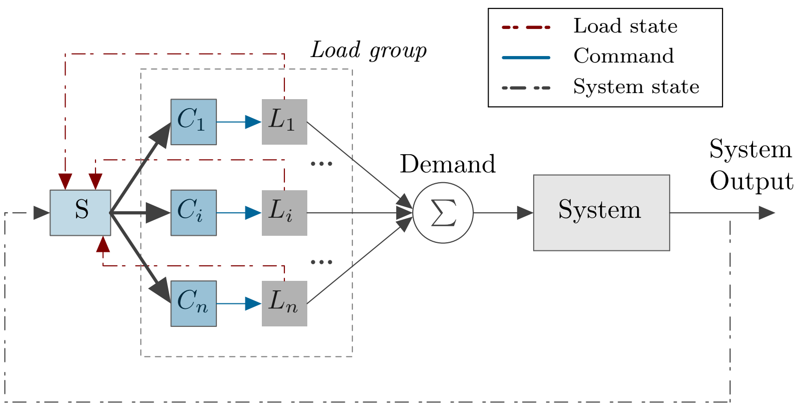
A system whose elements react based on local information to reach a system-level objective is defined as a decentralized system. System stability is guaranteed if the system-level state can be derived from local information [151]. Strictly speaking, the autonomous setup define above is a decentralized control setup. Our objective is to insist on the fact that no communication exists between the load and the central entity.

In the context of power systems, primary frequency control (frequency containment reserve, or FCR) is designed in a decentralized fashion [52]. Participating units react solely based on the system's frequency deviation from nominal. Deviations are measured locally and induce a proportional negative feedback response from the FCR providers.

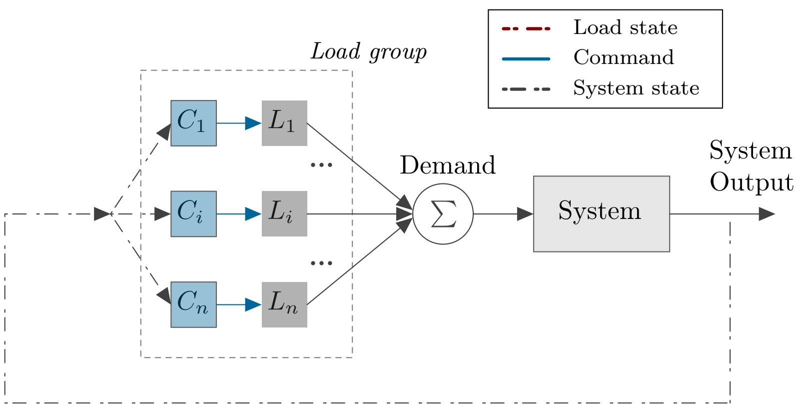
We also denote *autonomous control with group-level dynamics* the control settings in which local information is sufficient to reconstruct an image of the group-level state (e.g., group demand). This allows load to control the dynamics of their group response (e.g., oscillatory, etc.).



(a) Centralized. (C) : central controller. (L_i) Loads, $\forall i = 1..n$.



(b) Decentralized. (S) : broadcast center. (L_i, C_i) Loads & Local controllers $\forall i = 1..n$.



(c) Autonomous.(L_i, C_i) Loads & Local controllers $\forall i = 1..n$.

Figure 3.8: Possible load control setups.

3.3 TCLs in Frequency containment reserve

In this section, we focus on some interesting elements of the literature focusing on the autonomous control of TCL in primary frequency control.

It seems rational to exploit TCL within frequency control due to their inherent storage capacity. Their temperature may easily be changed with minimal impact on the end-user or related process. However, a large group of TCL does not necessarily react as as expected to frequency changes. Indeed, external control performed on a group of TCLs impacts demand *diversity* [32]. While loads respond to the unique frequency signal they locally measure, their consumption gets coordinated. The synchronizing effect this has on the group demand together with the load energy constraint induce a counteracting demand change known as the (energy) rebound, or rebound effect.

The rebound effect linked to load synchronization has been observed since decades in other contexts. It is described in [61] as load pickup. When applied to cold loads (refrigeration, the cold load pickup is the *demand when service has been restored after a prolonged outage* [95]. During the outage, cold loads got *synchronized*, as they are all stopped at the same time. Their regulated temperature gets progressively out of its authorized band. As soon as power is back on, they will all start consuming at the same time until the desired temperature is reached as represented on Fig.3.9.

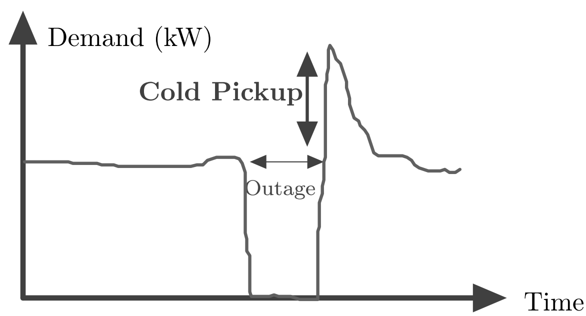


Figure 3.9: Load pickup following a power outage (illustration) [95].

Such group-level dynamics show up when TCLs are exploited in frequency response. In autonomous control setups, the local controller is the unique element capable of managing these dynamics. The challenge is to design a controller that will limit the impact of rebounds on the system, while only relying on local information. Centralized and decentralized control setups have a more direct way of counteracting complex dynamics thanks to their communication channel.

3.3.1 Demand control by adjusting temperature limits

The power demand of a group of TCL can be controlled by adjusting the limits of the interval in which the inner mass temperature of TCLs should remain.

In the 1980s, F. Schweppe *et al.* [126] proposes the famous FAPER control concept (i.e., Frequency Adaptive Power Energy re-Scheduler [125]), a form of autonomous FCR controller for small TCLs. Actually, Schweppe's main intention was to discuss the interest of spot pricing of electricity able to control power demand in a decentralized framework. In his view, it was crucial to operate a paradigm shift from the *generation-adapting-to-demand* to what's denoted as *homeostatic utility control*. In this new paradigm both generation and demand would adapt to each other through the use of communication and price signals.

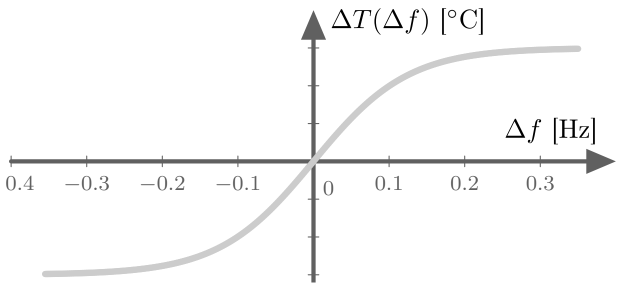


Figure 3.10: Temperature set-point adjustments to frequency deviation (FAPER).

The FAPER control is the load-side element of homeostatic control for frequency response. It considers that load could locally measure frequency deviations (from nominal) Δf and adapt their power consumption accordingly. The concept is applied to Thermostatically Controlled Loads (TCL) adjusting their temperature set-points by an amount $\Delta T(\Delta f)$ in response to frequency deviations (Table 3.3). The FAPER concept considered some saturation effect of the provided response (Fig.3.10).

Similarly to what is proposed by FAPER, the temperature limits of a cooling appliance are modified linearly with the frequency deviation (eq.

Table 3.3: High and Low temperature set point adjustments (FAPER) [126]

	$\Delta f < 0$	$\Delta f > 0$
T_H	$T_{max} + \Delta T(\Delta f)$	T_{max}
T_L	T_{min}	$T_{min} + \Delta T(\Delta f)$

(3.1) and (3.2)) in [127]. For instance, in case frequency decreases, the lower temperature limit increases and loads are switched-off.

$$T_H(t) = T_{H,0} - K_T \Delta f(t) \tag{3.1}$$

$$T_L(t) = T_{L,0} - K_T \Delta f(t) \tag{3.2}$$

As expected, the energy that is momentarily not consumed by the loads is recovered later on. The demand of the group increases compared to its initial level (rebound).

3.3.2 Stochastic temperature adjustments

Some other switching strategies aim at coordinating loads across the whole temperature state space, instead of focusing on the limits of the temperature interval. Such strategy actually lowers the rebound magnitude.

In [9], the thermal loads switching behavior (i.e., changing from on to off state and inversely) is randomly controlled. Instead of using temperature sensors to govern the moment at which load should run or stop, the load exploits switching probabilities λ_{OFF} and λ_{ON}. When a load is running, it has a probability λ_{OFF} to switch-off. The load then perform random trials at every time-step in order to decide on its state on the following time step. The two ON and OFF states as well as related switching probability define two-state Markov chain represented on Fig. 3.11.

The number of time-steps during which the load is ON of OFF are a random variables. The expected temperature level of the inner mass as well as its variance are dependent on the values of these variables. Furthermore, there exist two strict boundary temperature T_{ON} and T_{OFF}. The inner mass should not go out of the interval $[T_{OFF}, T_{ON}]$ (cooling appliance). Therefore, hard constraints are enforced to respect the extreme temperature limits. If the transition probabilities are chosen adequately, the probability for a load to reach the extreme limits is low, and the effect of hard constraints goes unnoticed.

The contribution of this study is to make the switching probabilities

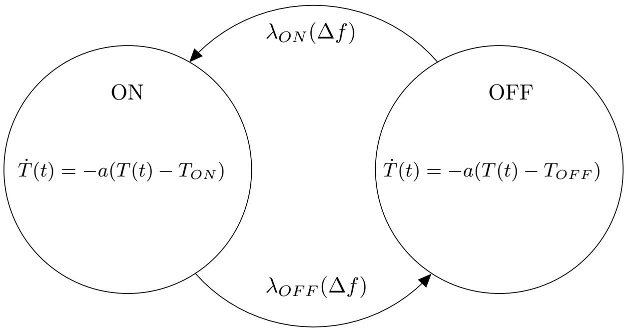


Figure 3.11: Markov chain representation of On and Off switching behavior [9].

dependent on the measured frequency deviation. The paper analyses the dynamical influence on the group demand.

3.3.3 Explicit integration of group-level dynamics in local decisions

A similar objective is pursued in [159] and [24]. These papers rely on a specific model previously used in [81], that we introduce below.

Centralized and decentralized approaches

In [81], the authors exploit a mixed-type modeling framework: an aggregate model is used to easily define control inputs that should be send to the individually modeled loads. The control setup is either centralized or decentralized (several test cases).

The core idea of the aggregate model is to cluster loads into N_{Bin} bins representing their run state (On or Off) and the (discretized) temperature interval in which they lie. A bin i contains a number of loads $n^i(t)$ at time t. The state-space (bins) is illustrated on Fig.3.12. Let's denote by $\mathbf{n}(t) = \{n^1(t), .., n^{N_{Bin}}(t)\}$ the number of load in each bin. The number of loads in the ON state correspond to the total demand of the group.

In [81] (with more details in [104]), the authors *estimate* the short term

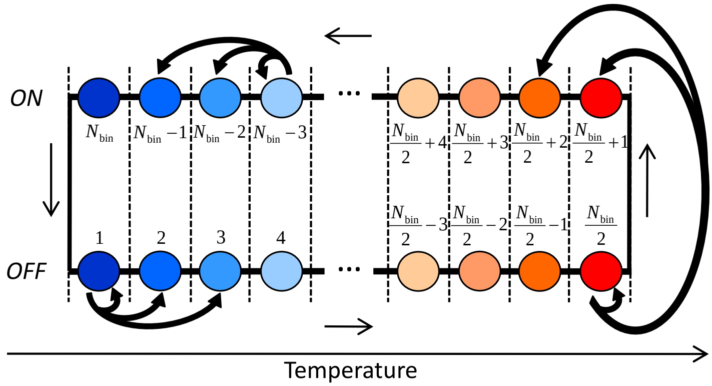


Figure 3.12: Discrete State Space used for a cooling TCL aggregate model [81].

evolution of the TCLs distribution $\hat{\mathbf{n}}(t)$ by deriving a linear model.

$$\hat{\mathbf{n}}(t+1) = A\hat{\mathbf{n}}(t) + B\mathbf{u}(t) \tag{3.3}$$

$$\hat{n}^{ON}(t) = \sum_{i=N_{bin}/2+1}^{N_{bin}} \hat{n}^i(t) \tag{3.4}$$

$$\hat{D}(t) = \bar{P}\hat{n}^{ON}(t) \tag{3.5}$$

The transition matrix A describes how loads $n^i(t)$ in bin i will transit to adjacent bins as a consequence of their natural behavior, while matrix B captures the impact of external commands $\mathbf{u}(t) = \{u^1(t), .., u^{N_{bin}}(t)\}$, which correspond to switching orders from the central controller. The total power demand of the group $D(t)$ is derived by counting down the number of loads in the ON state $n^{ON}(t)$ at time t and multiply it by the average power of the loads $\bar{P}$. Let's note that in practice, this power is not necessarily constant with time (parameter distribution, etc.).

In the decentralized setting, the mass distribution across the different bins is influenced by sending probabilities $u^i(t)$ that each load exploits to randomly decide on its state at the next time step. Model Predictive Control techniques are exploited to find the optimal commands $u^i(t)$. The demand $D(t)$ is controlled in order to follow a certain reference signal $r(t)$.

Using this model, authors in [104, 81, 103] have observed important elements.

- Loads synchronization leads to oscillatory behavior of the group's demand that must be counteracted.

- The service in which load participate (e.g., FCR) defines the reference signal $r(t)$ that must be followed by the group demand. The control must appropriately decide on the portion of the group's demand that can be offered as flexible capacity for external control. If this portion is too large, the loads are likely to go out of their temperature limit. If it is too small, an opportunity is lost. Some advanced energy management system are required to balance control performances with local temperature impact.

- The per load benefit of participating in frequency control is small (also noted in [88]). This strongly limits the attractiveness of small load control on the end-user side.

More complex TCL models taking into account second-order dynamics are studied using a similar decentralized control framework in [158].

Autonomous control

Relying on these conclusions and modeling framework, [159] and [24] discuss possible ways to compute the control inputs in an autonomous fashion. Autonomously, each load $j \in \{1, .., n_{TCL}\}$ in the group computes at each time step a switching probability $q_j(t) \in [0, 1]$. The load will switch state by comparing the $q_j(t)$ to another number randomly that is randomly chosen between 0 and 1.

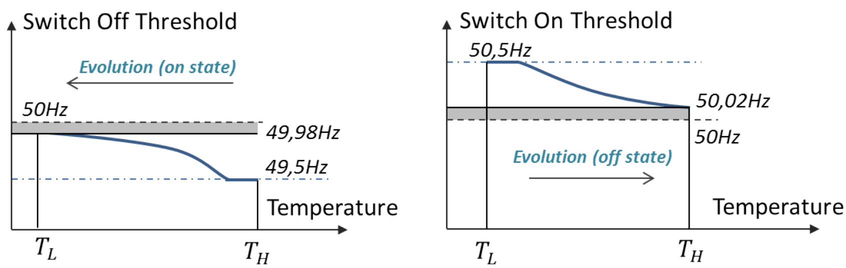
The successive $q_j(t)$ are function of both load-side elements as well as local estimate of the group-level states. Simply based on local frequency measurement and knowledge about the individual behavior, each load is able to estimate the group's state autonomously. In [159], additional constraints are introduced: a variable consumption profile (Startup dynamics) as well as lockout constraints. The lockout constraint is the minimum time immediately following a state transition during which no other transition can occur. In [159], each loads is able to keep track of the number of locked-out loads and adapt its own switching decision accordingly.

3.3.4 Pilot Projects

In [146], a UK-based pilot project focuses on the use of small refrigerator within primary frequency control. The above described references [127] and [9] are two different theoretical descriptions of the system used in the project. All running fridges (i.e., on state) autonomously switch off in case frequency decreases below a defined threshold. Inversely, loads will anticipate their start if they measure sufficiently large positive frequency deviation.

The switching frequency thresholds are randomly set by the load around a reference value. The random thresholds are chosen in order to avoid large load synchronization .

A reference threshold is defined and common to all loads in the group. It is illustrated on Fig.3.13. It evolves according to the load's run state and the temperature. The shape of the reference threshold minimizes the chances of observing successive switching. For instance, a load that has just switched to the ON state has a very low probability to switch OFF immediately after.



(a) On State (switch to off state) (b) Off State (switch to on state)

Figure 3.13: Reference Switching Frequency threshold (illustrative) [146].

At each time they switch state, the different loads will select a random threshold around this reference switching profile. By comparing their own threshold to the frequency they measure, loads will react to extreme frequency changes and support the system. The random feature allows to spread the efforts smoothly.

The performance of the group and its ability to provide a proportional response to frequency deviations is not clearly assessed in [146].

Experimental results of central control performed on 25 fridges are described on [87]. Similarly, 26 refrigerators and space-heater, a waste-water treatment plant and 10 general relay-controlled loads are exploited in [44]. Different local control law are applied and leading to good control performances. The long term impact of the involved control schemes is however unclear.

3.4 ECLs in Frequency Containment Reserves

The literature focusing on the ECL type of loads (mainly: electric vehicle, batch water heaters) has not been very extensively discussing how autonomous control could take place. However, many studies look at centralized and decentralized control frameworks (Table 3.1). We discuss briefly below interesting elements about the following points.

1. Aggregate models development.
2. Consumption scheduling: discussions on the optimal way to schedule the consumption of ECL along a day in order to minimize energy purchase costs and maximize the aggregate flexible capacity provided as FCR in a decentralized fashion.
3. Centralized control: gave inspiration for autonomous control.
4. Elements on autonomous control.
5. Pilot Project outcomes.

In general, studies focus on relatively short-term simulations. The performance of the proposed control framework are usually evaluated in realistic simulations but that are not spanning on more than few hours. The economic performances of the proposed control scheme are therefore rarely assessed. In [7], authors look at the profitability of plug-in electric vehicles offering different types of frequency support services (FCR, aFRR and RR). The results are very optimistic (up to 100€per month and per load!), but rely on a different assumption than ours: the batteries of the electric vehicles offer a certain portion of their storage capacity to the grid operator that they make available for charging and discharging purpose at each time where they are actually connected to the grid.

3.4.1 Developing Aggregate models

Several propositions exist in the literature about the construction of aggregate models. In [3, 2], authors based their developments on clustering techniques, Markov chains and queuing theory. The objective is to efficiently compute the aggregate consumption of a group of loads based on reference consumption profile, and on their probability of occurrence. Model Predictive Control techniques are then applied for controlling this aggregate consumption. The computational burden is very well diminished. Yet, the system perspective and load impact of the control is absent from the discussions.

A similar approach is taken in [73], where the average static droop provided by a group of electric vehicles is evaluated. This droop depends on the batteries state-of-charge, or SOC. For instance, batteries that are fully charged cannot consume more energy. The distribution of the load SOC and its evolution with time are used to estimate the aggregate capacity that the group could instantaneously deploy.

However, only very short run simulations are undertaken. The consequence of constant service provision on the SOC (feedback loop), and therefore on the offered capacity are not considered.

3.4.2 Scheduling load consumption to insure FCR capacity

An important topic lies in deciding on optimal time a flexible appliance should consume energy. Many papers discuss this problem, but some of them specifically address ECL and a combined energy-flexibility objective. Indeed, load schedule is decided to minimize the overall energy price as well as guaranteeing a certain amount of aggregate flexible capacity. This capacity is later exploited within frequency control.

This combined scheduling problem is e.g., discussed in [64]. The authors in [149] compare the centralized and decentralized scheduling of the amount of capacity that should be delivered at each hour. Centralized scheduling means that loads submit their preferences (wilingness to pay, bid) to a central entity that later decides on the scheduling. Decentralized is based on price. The detailed information about the different loads is not accessible to the grid operator. However, it can send prices and collect aggregate demand of the group of loads. The optimal price is thereby decided on an iterative manner. Such type of scheduling is specifically useful to cope with data privacy issues.

In [43], such scheduling approach is performed on a stochastic way and take uncertainty into account.

3.4.3 Centralized control of batch water heaters

In [84], the authors discuss the use of batch water heaters for frequency control in central control framework. The considered water heaters have a fixed power rate p. They require some predefined amount of energy E to be delivered before a certain time deadline t_e. At each time, the load knows its own energy state $0 < e(t) < E$, that should reach E for $t > t_e$. The model is therefore very similar to electric battery charging.

The loads are able to estimate at every time *how far they are* from reaching their energy requirement. They translate their estimation into a simple factor $\gamma(t) = \frac{E-e(t)}{p(t_e-t)}$ that must stay below 1. Indeed, when γ=1, the load is forced to run. The load can postpone its energy consumption as long as γ is below 1. This factor is called the remaining power consumption ratio.

The system operators sends two *thresholds* values $\gamma_{on}(t)$ (closer from 1) and $\gamma_{off}(t)$ (closer from zero) to the water heaters. Each load i compares them to its own $\gamma_i(t)$. All loads which ratio is above the γ_{on} threshold should be running. All loads which ratio is below the γ_{off} threshold should be stopped. If the system requires all loads to stop, it sends $\gamma_{off} = 1$.

Though this strategy requires the use of centralized information, the same kind of principle could be applied to autonomous controllers. The threshold would simply need to be defined according to the measured frequency deviations.

3.4.4 Autonomous control of ECLs for FCR provision

Autonomous Electric vehicle charging

A large number of studies have explored the use of electric vehicles (EV) charging delivering flexibility. Yet, no many of them explore autonomous control solutions. The autonomous charging of EVs combined with frequency control participation is considered in [98].

The aim of the article is to determine a *local* control logic able to either charge electric vehicles or maintain its state of charge adapting the power consumption to provide frequency response. It is therefore what we refer to as an *autonomous* control framework (it is characterized as *decentralized* in [98]).

The main idea is simple: loads will provide a proportional response to frequency deviations with a *droop* $K(e(t))$ (gain) that differs according to the state-of-charge (energy state) of the load $e(t)$. This droop must be as large as possible in order to offer the largest achievable flexibility. However, it is adapted respect the load's own constraint (in charge mode or not, etc.).

The implemented control scheme is tested on several simulated vehicles in a 10h simulations. This is sufficient to test the effectiveness of the control scheme in the short-run. However, the study does not discuss the impact on the overall system neither economic considerations. Similar developments can be found in [45].

Heating appliances

Dishwashers are controlled in an autonomous way to react to large frequency excursions in [144]. The dishwasher profile is problematic. Indeed, it is composed of two successive heating phases where water is heated up before it can be used in the washing phase. The two heating phases require the use of heating coils consuming a relatively large amount of power. They are separated from each other by a low-power washing phase, illustrated on figure 3.14.

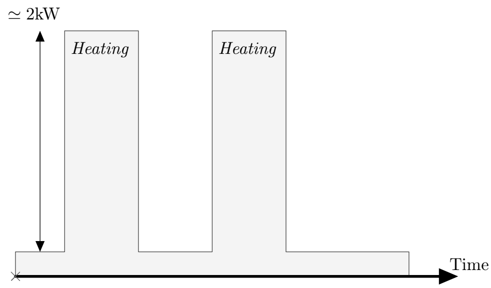


Figure 3.14: Illustration of a Dishwasher profile (inspired from [91]).

The paper assumes that dishwashers can be rapidly switched on an off during their consumption cycle. The FCR controller decides to suddenly stop the dishwashers if the frequency goes beyond a certain threshold (under-frequency load shedding). The main contribution of the paper is to analyze how dishwashers are progressively switched back on. Imposing random delays before the load can resume consumption leads to good performances. This is an example highlighting how the random nature of autonomous controllers can be useful. Random numbers are able to smooth out the negative consequences of imposing a sudden coordination to loads leading to a loss in demand diversity and strong rebound effects. The random delay also allows to average out the consumption profile of the involved loads.

Pool pumps (and beyond)

The control philosophy developed in [84] is developed in [108] in an autonomous fashion. A so-called *universal command* is developed, to which loads adapt and randomly set their on/off state accordingly. Their decision depend on how much they where started/stopped the same day, knowing their needs to run for a certain amount of time.

The paper observes the dynamics of a controlled group of pool pumps, though the controller is usable for many other appliances (other ECLs). The study mentions future research ambitions: taking a *system operator perspective* and derive costs and characterize the flexible capacity provided at group-level.

3.4.5 Pilot Projects with ECL

Probably the most famous pilot project about ECL, the GridWise pilot project, is described in [66]. The objective of this project was to provide frequency support in emergency situations by intelligently curtailing small ECL, such as water heaters. Some interesting considerations about implementation issues and cost components are to be found in the project report, and will be exploited in the following chapters of this work.

Several other projects have been conducted on the use of ECLs in system balancing (e.g., vehicle-to-grid [76]). The Linear Project [97] studied the use of batch water heaters, electric vehicles and wet appliances in different context. Among the explored business cases, the *balancing* business case could be considered quite close to the use of ECL within FCR. Some interesting elements about the cost effectiveness and potential benefits are discussed. Loads would diminish at best 5-10% of their annual energy costs by providing flexibility services.

Finally, let's mention the Swiss2grid project [122]. In a local neighborhood, load management (EV charging, water and space heating, washing machine and dishwasher) is implemented for voltage control and peak load decrease (transformer life time). The interesting contribution of the followed approach lies in the communication infrastructure that is used. Indeed, the project has looked at different ways to coordinate loads: full two-way communication between the loads and a central entity, or local exchange information with direct neighbors. It appears that local communication is enough reach performances that are very close to the ideal case where all information is transferred to a central entity.

3.5 State-of-the-art conclusions and contributions

Here are the principal elements of conclusions that can be extracted from the existing literature covering the use of small loads for FCR provision, and autonomous control frameworks.

1. *Controller design.* The problem of controller design is to define adequate control laws able to influence the group demand, make it frequency responsive and manage rebound errors (group level dynamics). These tasks must be performed only relying on local information (autonomous control).

2. *Aggregate models.* In the literature about Energy Constrained Loads (focus of the following chapter), we could not find any study focusing on the design of autonomous controller and, at the same time, tracking the evolution of the controlled demand by using aggregate models, in the aim of delivering frequency containment reserves. This is therefore one of the core elements of our next chapters.

3. *Control performances in the short-run.* When control performances are assessed, the paper usually restricts their analyses to short-term simulations. However, the dynamics of the demand of ECL groups are such that long run simulations are essential to capture the effect of energy rebound (energy constraints). Studies focusing on the short run simulations do not capture the rebound problem adequately. This is therefore a second important element that will be developed in this work.

4. *Costs and Benefits.* The benefits of exploiting ECL within FCR are discussed in a few studies. Their results are however either overly optimistic, or do not analyze the overall system consequences (frequency quality in the long-run, etc.).

In short, the use of Energy Constrained Load as Frequency Containment Provider is rarely discussed. Furthermore, the perspective of the system operator is not sufficiently represented. Furthermore, the impact of energy rebound is usually overlooked. As we will show in the rest of this work, introducing autonomous control within FCR has, for some loads, an overall positive economic impact. In addition, this economic impact is influenced by long-term trends, that are usually not discussed in the literature. Our work seems therefore very well justified.

What are the exact control performances of the group in a realistic context ? In the short-run (frequency stability) ? And, in the long-run (frequency quality) ? What is the overall system impact of ECL control on the reserve provision process ? Will the system gain installing small controllers on every load? What are the characteristic of the ideal ECL ? These are the questions that will be discussed throughout the rest of this work.

Chapter 4

Aggregate behavior of Energy Constrained Loads

Chapter summary

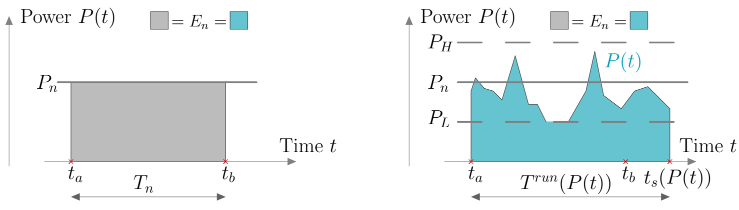
In the beginning, there was nothing but Chaos, out of which emerged spontaneously everything that exists.

Hesiod, *Theogonia* c.700 BCE (adapted).

Let's imagine that a very large number of electric appliances are equipped with on-board controllers. These controllers can change the power at which the appliance consumes its energy. There are several typical examples: the load can deffer its starting time, the load may suddenly be stopped and started back later on, or the active power it consumes may be set freely within a certain range thanks to power electronics equipment. But what are the consequences on the group-level demand of using such local controllers? What if such controllers take decisions autonomously ?

In the rest of this work, we will focus on a specific type of electrical appliance: Energy Constrained Loads. The first three sections of this chapter study large groups of ECL which are left uncontrolled. Our objective is to describe the dynamics of the group-level demand when ECLs of multiple types (i.e., with different power rate, etc.) are aggregated, and when the individual starting times are random. Indeed, the instant at which an electric appliance is plugged into the network by the end-user can be considered as a random variable. Consequently, the group-level demand will experience some volatility around its expected level. This volatility should be small enough such that the control reactions imposed to the group can be distinguished from its natural variable behavior.

Our analyses end with clear-cut observations: only a massive implementation program, in which several hundreds of thousands of loads are involved, can guarantee sufficiently good performances.



(a) Without control. (b) With Power control

Figure 4.1: Energy Constrained Load model.

An ECL has the following features, illustrated on Figure 4.1.

1. An arrival time t_a from which the load starts consuming energy.
2. A controllable power rate $P(t) \in [P_L, P_H]$.
3. A natural power level P_n at which the load consumes energy when uncontrolled.
4. A fixed energy need E_n (energy constraint) requested by the ECL's user.
5. A user-defined time deadline T_{dl}. The required energy E_n must be consumed before this time deadline, that is in the interval $[t_a, t_a + T_{dl}]$.
6. A natural run time $T_n = E_n/P_n < T_{dl}$: the necessary time for the load to consume E_n when running at its natural power rate P_n (no power control).
7. A variable run time $T^{run}(P(t))$ that depends on the controlled power $P(t)$.

4.1 Load arrivals in large ECL groups

The instant at which a user requests an appliance to start defines the load's *arrival*. The number of loads arriving in a group at a certain time t needs to be characterized. In practice, the user's behavior follows a random pattern.

The function $C(t)$ is a *counting process*, accounting for the number of loads that have arrived in the time interval $[0, t)$. It is a stair-shaped function of which each step represents a load arriving in the group. Each step

i (a load arrival) is situated in time by a random sequence of *epochs* $\{Y_i\}$. Each Y_i corresponds to a load arrival time t_a. The time that separates two successive arrivals is called the inter-arrival interval X_i.

The starting density $S(t)$ is the time derivative of the counting process. It is a condensed version of all information about the different load starts. There are multiple possible representation of the starting density, both in discrete S_k and continuous-time $S(t)$ modeling framework.

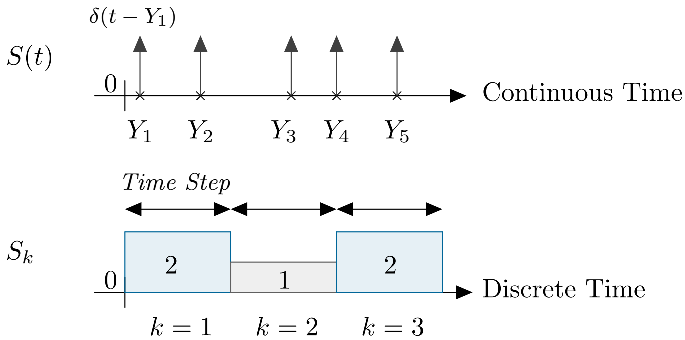


Figure 4.2: Two equivalent representations of the starting density. Top: Continuous-time model. Bottom: Discrete-time model.

These are equivalent and represented on Figure 4.2. The unit of $S(t)$ corresponds to the inverse of the selected unit of time. It is a sequence of Dirac delta's separated by random time intervals. The discrete time version is dimensionless and corresponds to the number of loads that have arrived withing a certain time step of length Δt. The number of arrived loads in the different time steps are random variables with known distribution.

4.1.1 Counting down arrivals with Renewal processes.

The starting density $S(t)$ (or S_k) and the counting process $C(t)$ are random functions called renewal processes [138]. The main assumption defining renewal processes is that the rate at which arrivals occur stays independent from previous arrivals (i.e., memoryless nature). A typical example: arrivals of customers in a queue (queuing theory).

> A renewal (counting) process $\{C(t), t \leq 0\}$ is a non-negative integer-valued stochastic process that registers the successive occurrences of an event during the time interval $[0, t)$, where the times between consecutive events are positive, independent,

identically distributed random variables. (Slightly modified from [138])

The number of appliances that request to start at some point in time may depend on exogenous elements: time of day, external temperature, etc. At the condition that these parameters are known at time t, the future evolution of load arrivals are independent of their past. This conditional independence is known as the Markov property. Load arrivals can therefore be modeled by renewal processes.

4.1.2 An intuitive example: the Binomial Process

As a start, let's consider the discrete representation of time with time-steps $k \in \{0, 1, .., K\}$ (i.e., time starts at $k = 0$) of length Δt and time t,the continuous representation of time. Let's consider a large group of nl loads. Loads are initially all stopped (or *idled*). Each load is associated with a certain probability p_k to arrive at a certain time step k. This probability is very small as loads run relatively rarely w.r.t. the time they spend being idled. Let's assume that the different load arrivals occur independently from each other and S_k be the number of loads arriving at time step k.

In such case, the successive S_k are random variables that follow a binomial distribution. Below, ${}^aC_b$ is the number of b-combinations chosen among a elements (without repetition). Also, $Pr\{Z = z\}$ stands for the probability of the random variable Z to be equal to some value z.

$$S_k \sim B(nl, \mathrm{p}_k) \qquad \Leftrightarrow \qquad Pr\{S_k = z\} = {}^{nl}C_z \, \mathrm{p}_k^z \, (1 - \mathrm{p}_k)^{nl-z} \quad (4.1)$$

Three binomial processes are illustrated on Figure 4.3. As can be observed, similar behaviors emerge from different groups in case the product $nl\mathrm{p}$ is the same. As this product gets higher, the distribution of arrivals tends toward a normal distribution. Furthermore, as this product grows, the distribution variance gets smaller relatively to its mean.

4.1.3 The law of rare events and the Poisson process

As both nl is large and p_k us very small, the Binomial process may be conveniently represented by a Poisson process. As shown on Fig.4.4, the static probability distributions match almost perfectly. This approximation is known as the law of rare events [138]. In a continuous-time framework, the law of rare events boils down to impose that multiple events have zero probability to occur exactly at the same time.

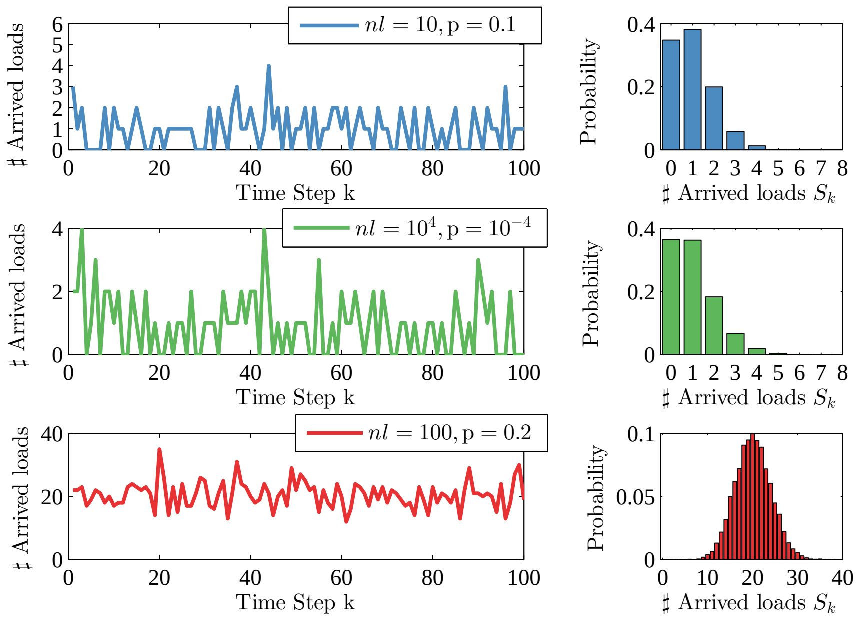


Figure 4.3: Binomial processes for different parameter values. Left charts: successive realizations on 100 time steps. Right charts: histograms of the corresponding distributions (10^4 realizations).

Computationally speaking, Poisson processes are less demanding to simulate. As illustration, a hundred different sequences (Trials) of both Poisson and Binomial processes are evaluated. Each generated sequence of Poisson distributed variable is 10^5 time steps long while each binomial sequence is only 10^3 time steps long. As shown on Fig. 4.5, the Poisson trials went 10 times faster at generating 100 times more data than the Binomial trials. The Poisson distribution is approximately 10^3 faster than the Binomial evaluation.

The Poisson process is very often exploited in the literature. A Poisson process is a renewal process whose inter-arrival intervals are i.i.d. exponentially distributed [123].

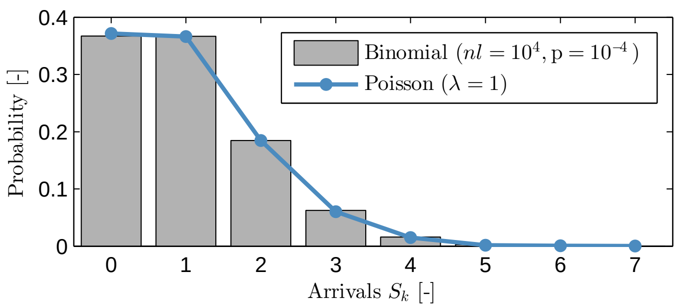


Figure 4.4: Probability distribution of two Poisson and Binomial random processes (Two sequences long of 10^4 time-steps).

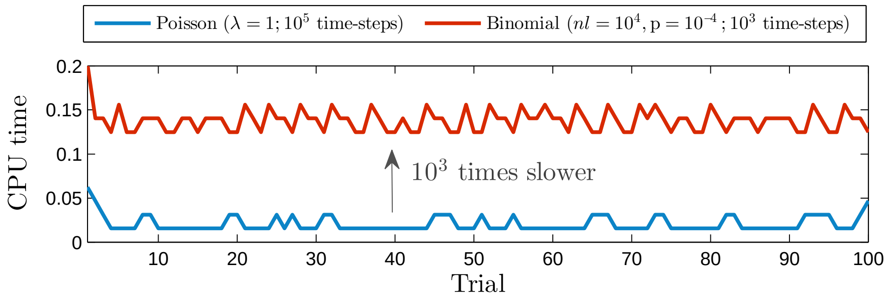


Figure 4.5: CPU times required to evaluate a sequence of Poisson and Binomial realizations. Theoretical mean and variance of each process are approximately all equal to 1.

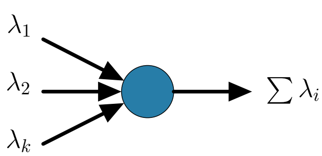
It is a continuous-time process with the following properties. Let X_i be the time intervals between the i^{th} arrival and its predecessor ($X_0 = 0$).

1. X_i's are i.i.d. exponentially distributed of parameter λ and density $f_X(x) = \lambda e^{-\lambda x}$.

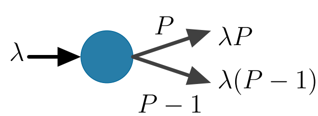
2. Orderliness: Inter-arrival intervals are strictly positive, $X_i > 0, \forall i$. Multiple events cannot occur at the same time.

3. Memorylessness: $Pr\{X > a + b\} = Pr\{X > a\}Pr\{X > b\}$.

4. Merging and Splitting[1] : the sum of two Poisson is Poisson and a Poisson to which some realization have a certain feature with known probability will split into two Poisson.



(a) Merging Property



(b) Splitting Property

The starting epoch of arrival i is $Y_i = \sum_{k=0}^{i} X_k$ and the Poisson counting process is $\{C(t) = k | (\sum_{i=0}^{k} X_i) < t\}$.

[1]see : `http://web.mit.edu/modiano/www/6.263/lec5-6.pdf`

In case the parameter λ (rate) is constant, the Poisson process is said to be homogeneous. In this case, the probability to count n arrivals in the interval $[0, t)$ is computed as follows.

$$Pr\{C(t) = n\} = \frac{(\lambda t)^n \exp(-\lambda t)}{n!} \tag{4.2}$$

In realistic situations, users preferably consume at specific hours in the day, seasons in a year. Therefore, the rate $\lambda(t)$ at which loads arrive is time varying. Poisson processes with variable rate are called *non-homogeneous* and are easily found in reality. For example, the variable arrival rate of customers to a fast food restaurant is represented on Fig.4.7.

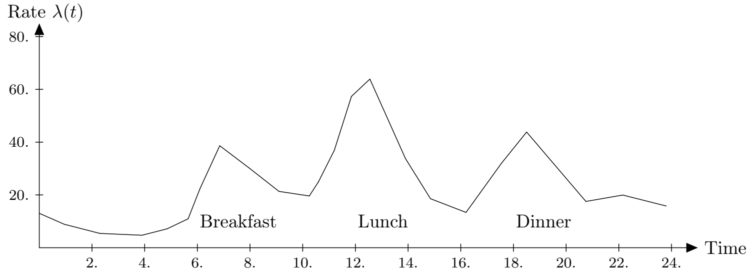


Figure 4.7: Customer arrivals to a fast food restaurant [94].

We use the following notations adapted from [94] and [123].

Notation

t	Time.
$C(t)$	Number of events by time t.
$\lambda(t)$	Instantaneous arrival rate at time t (intensity function).
$m(t)$	Cumulative intensity function.

Let's note that non-homogeneous Poisson process are renewal processes (i.e., i.i.d. inter-arrival intervals). Indeed, inter-arrival intervals are identically distributed and conditionally independent. Their distribution is *conditional* (i.e., depends on) to a time-varying parameter $\lambda(t)$. In this case, the inter-arrival intervals distribution loose their stationary nature.

In the fast-food restaurant example, the counting process $C(t)$ simply counts down the amount of customers that have arrived in the open interval $[0, t)$. In order to account for the time-varying nature of the intensity function $\lambda(t)$ (rate), it mus be integrated in the c*umulative intensity function*

$m(t)$, which is used to characterize $C(t)$.

$$m(t) = \int_0^t \lambda(s)\, ds \tag{4.3}$$

$$Pr\{C(t) = n\} = \frac{m(t)^n \exp(-m(t))}{n!} \tag{4.4}$$

In such case, the expected value of the counting process at time t is $\mathbb{E}[C(t)] = m(t)$.

Equivalent discrete-time formulation.

In the context of this work, it is legitimate to assume that multiple events cannot occur simultaneously. In a continuous-time modeling framework, successive load arrivals will be modeled using a counting Poisson process $C(t)$. Alternatively, we may prefer to use its time derivative $S(t)$. Let time t be expressed in seconds $[s]$. The discrete-time equivalent S_k is dimensionless and is obtained evaluating the continuous counting process at each time-step k of length Δt.

$$S(t) = \frac{d}{dt}C(t) = \sum_{i=0}^{+\infty} \delta(t - Y_i) \qquad [s^{-1}] \tag{4.5}$$

$$S_k = C(k\Delta t) - C(\Delta t(k-1)) \qquad [-] \tag{4.6}$$

The number of arrivals within a time-step are Poisson distributed. Let's define $\tilde{m}$ as the cumulative intensity function up to current time step [123].

$$\tilde{m}(t, \tau) = \int_0^\tau \lambda(t + s)\, ds \tag{4.7}$$

$$Pr\{S_k = n\} = \frac{\tilde{m}(k\Delta t, \Delta t)^n \exp(-\tilde{m}(k\Delta t, \Delta t))}{n!} \tag{4.8}$$

The integer valued sequence $(S_1, S_2, .., S_n)$ is a sequence of Poisson distributed random variables of parameter $\lambda_k = \tilde{m}(k\Delta t, \Delta t)$. For homogeneous Poisson processes, $\tilde{m} = \lambda\Delta t$. Poisson realizations are shown on Fig. 4.8 for different λ.

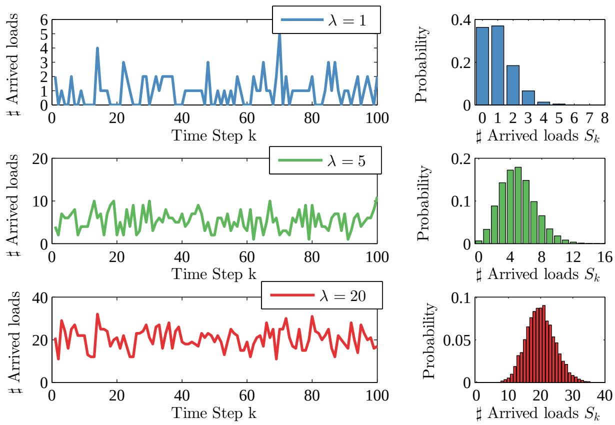


Figure 4.8: Homogeneous Poisson processes for different parameter values (discrete-time equivalent). Left charts: successive realizations (100 time-steps). Right charts: Histograms of the corresponding distributions (10^4 realizations).

4.2 The homogeneous ECL group.

Let's consider an *homogeneous* group of ECL whose arrivals are described by Poisson process. In an *homogeneous* group, ECLs share the same parameters [104]. Loads are left uncontrolled and run at their constant natural power rate P_n (kW). The power demand of the group $D(t)$ (kW) at time t is the product of the number of running loads $N(t)$ (dimensionless) with their power P_n (kW).

$$D(t) = N(t)\, P_n \qquad [kW] \qquad \text{uncontrolled case} \tag{4.9}$$

As a consequence of their fixed energy needs E_n, each load stops exactly $T_n = E_n/P_n$ instants after it has arrived. Therefore, the number of loads $N(t)$ running at time t can be computed by counting down the arrivals that occurred in the T_n previous instants, as all loads that have arrived earlier in time are stopped. According to (4.6), this gives

$$\begin{aligned} N(t) &= \int_0^{T_n} S(t-s)\, ds \qquad [-] \qquad \text{uncontrolled case} && (4.10)\\ &= C(t) - C(t - T_n) && (4.11) \end{aligned}$$

The number of running loads is a random process with strong autocorrelation. Indeed, its value at current time depends on its value in the T_n previous time steps. In other words, the time delay T_n in (4.10) and (4.11) engender an infinite dimensional problem [109].

As shown in [113], the number of running loads $N(t)$ is a Poisson random variable of parameter $\tilde{m}(t, T_n)$. In the Poisson distribution, the mean and variance are equal to the unique distribution parameter (dimensionless).

$$\tilde{m}(t, T_n) = \mathbb{E}[N(t)] = \mathrm{Var}(N(t)) = \int_0^{T_n} \lambda(t-s)\, ds \qquad [-] \tag{4.12}$$

Thereby, we may compute the demand's expected value and variance.

$$\mathbb{E}[D(t)] = \tilde{m}(t, T)P_n \quad \& \quad \mathrm{Var}[D(t)] = \mathrm{Var}[P_n N(t)] = (P_n)^2 \tilde{m}(t, T_n) \tag{4.13}$$

In case of constant rate λ, the formulation is simpler.

$$\begin{aligned} \mathbb{E}[D(t)] &= \lambda T_n P_n = \lambda E_n \qquad [kW] && (4.14)\\ \mathrm{Var}[D(t)] &= \mathrm{Var}[P_n N(t)] = P_n^2 \lambda T_n \qquad [kW^2] && (4.15) \end{aligned}$$

The demand $D(t)$ is illustrated on figure 4.9. The zero-centered counting process $C_0(t) = C(t) - E[C(t)] = C(t) - \lambda t$ is shown. This process $C_0(t)$ is the integral of the arrivals that occur is excess of the average arrival rate. The difference between $C_0(t)$ and its delayed version $C_0(t-T)$ indicates the level of the actual $D(t)$ w.r.t. its expected value.

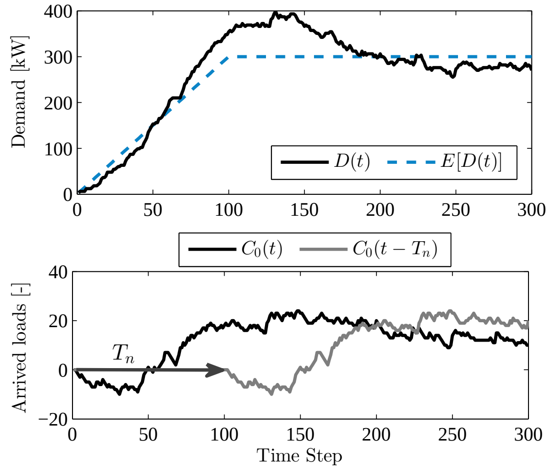


Figure 4.9: Power demand $D(t)$ (top) and zero-centered counting process $C_0(t)$ (bottom) of a large homogeneous group of ECL ($t = [0, 300]$). Parameters: $P_n = 3kW$, $\lambda = 1[s^{-1}]$, $T_n = 100[s]$.

4.3 Heterogeneous Group

Groups whose loads do not share identical parameters are heterogeneous. The time T_n and power P_n parameter are in general random variables. In this section, we study the influence of parameter variability on the group demand.

Let's denote the joint probability distribution $\Omega_{P,T}(t)$ as the *parameter distribution at arrival*. It evaluates the probability for a load *arrived at time* t to have parameters $P_n = P$ and $T_n = T$. In general, the distribution is time varying. It exists in the interval ($T \in [T_m, T_M]$; $P \in [P_m, P_M]$). That is, a load will run for no less than T_m and no more than T_M instants. A uniformly distributed $\Omega_{P,T}$ is shown on Fig. 4.10.

$$\int_{T_m}^{T_M} \int_{P_m}^{P_M} \Omega_{P,T}(t)\, dt = 1 \qquad \forall t \tag{4.16}$$

$$\Omega_{P,T}(t) = 0 \qquad \forall P, T \text{ s.t. } P < P_m \cup P > P_M \cup T < T_m \cup T > T_M \tag{4.17}$$

In practical cases, the information about the running loads (i.e., their parameters) is not necessarily accessible in real-time. It is therefore important to estimate the group demand based on the parameter distribution. The estimated demand $\hat{D}(t)$ corresponds to the expected value of $D(t)$. Let's

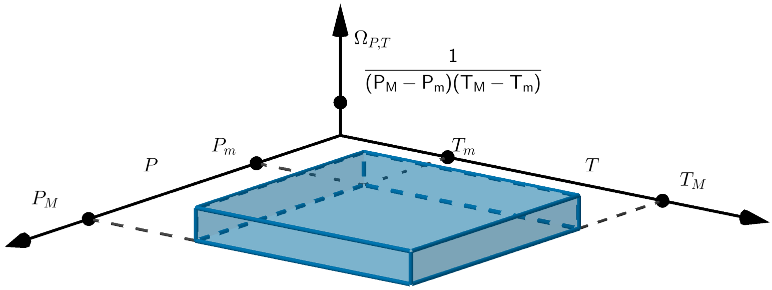


Figure 4.10: Parameter distribution at start $\Omega_{P,T}$. This illustrates a constant and uniform distribution.

define the time delay $\tau \geq 0$ that expresses the time interval between current time t and a previous instant $t - \tau$. As perceived from time t, a running load arrived at time $t-\tau$ has a probability $\Omega_{P,T\geq\tau}(t)$ to run at power rate P.

The expected demand $\hat{D}(t)$ of the group is derived below, where $S(t)$ is supposed to be known.

$$\hat{D}(t) = \int_0^{T_M} \int_\tau^{T_M} \int_{P_m}^{P_M} S(t-\tau)\; P\,\Omega_{P,T}(t-\tau)\,dP\,dT\,d\tau \tag{4.18}$$

The expected power demand $\hat{D}(t)$ is a convolution of the starting density $S(t)$ with the expected power rate of loads running longer than the time $t-\tau$ at which the start occurred.

4.3.1 The average load: expected power rate

At time t, the *expected power rate* $\omega(t,\tau)$ is the conditional expected value of the parameter power P given that the related load has started at time $t-\tau$ and must still be running on t. In other words, it is the conditional expected value of P given $T \geq \tau$.

$$\omega(t,\tau) = \mathbb{E}[P|T \geq \tau](t) = \begin{cases} \int_\tau^{T_M} \int_{P_m}^{P_M} P\,\Omega_{P,T}(t-\tau)\,dPdT, & \tau \in [T_m, T_M] \quad (4.19) \\ \int_{T_m}^{T_M} \int_{P_m}^{P_M} P\,\Omega_{P,T}(t-\tau)\,dPdT, & 0 \leq \tau < T_m \quad (4.20) \\ 0, & \text{elsewhere} \quad (4.21) \end{cases}$$

In the above expression, $\omega(t,\tau)$ can be understood as a power consumption profile. In case of constant distribution $\Omega_{P,T}$, this profile is constant with time t and varies along the time delay τ. On Figure 4.11, such profile is represented. As can be observed, it is constant in the interval $\tau \in [0, T_m]$ (eq. (4.20)). It then decreases progressively to zero, as the time delay approaches the maximum time duration T_M (eq. (4.19)).

Equation (4.18) is rewritten using the expected power rate.

$$\hat{D}(t) = \int_0^{T_M} S(t-\tau)\omega(t,\tau)d\tau \tag{4.22}$$

Assuming a constant distribution $\Omega_{P,T}$ and that the starting density can be represented by its mean value λ, let $\hat{D}_0$ be defined as the *baseline* consumption level. It has two equivalent formulations in this case.

$$\begin{aligned}
\hat{D}_0 &= \lambda \int_0^{T_M} \omega(\tau)\, d\tau & (4.23)\\
&= \lambda \int_{T_m}^{T_M} \int_{P_m}^{P_M} PT\Omega_{P,T}\, dP\, dT & (4.24)\\
&= \lambda E^{av} & (4.25)
\end{aligned}$$

In equation (4.25), E^{av} represents the energy content of the average power cycle $\omega(\tau)$. Let's note the parallel between (4.25) and (4.14).

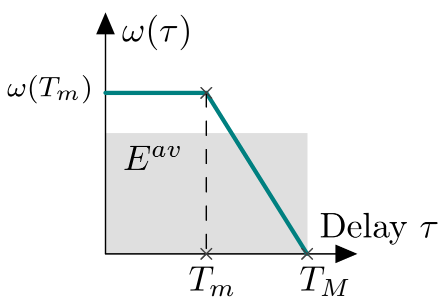


Figure 4.11: Expected power rate of load having started τ instants earlier. The illustrated case corresponds to uniformly distributed $\Omega_{P,T}$ (Fig. 4.10).

4.4 Transition from homogeneous to heterogeneous nature

Time varying parameters introduce complex dynamical behavior into the group's demand evolution. In order to illustrate this, let's consider an extreme example. A group of ECL is characterized by a constant starting density $S(t) = \lambda$ and the following elements.

- Before $t = t_1$, the group is homogeneous with power P_n and run time T_n.

- At time $t = t_1$, the group turns progressively heterogeneous. New loads arriving in the group have uniformly distributed run times $T \in [T_m, T_M]$. We select the limits symmetrically around T_n. This means that the expected run time is the same than in the homogeneous case : $0.5(T_m + T_M) = T_n$.

- From $t_2 >> t_1$, the group recovers progressively its homogeneous nature. New loads run for a fixed duration T_n.

As parameter $P = P_n$ is constant, the distribution $\Omega_T(t)$ is sufficient to describe the variability of the group. For compactness, we exploit the time-window function $U(t, t_a, t_b) = [\mathcal{H}(t - t_a) - \mathcal{H}(t - t_b)]$, where $\mathcal{H}$ is the Heaviside function.

$$\Omega_T(t, \tau) = \delta(\tau - T_n)\ [U(t, 0, t_1) + \mathcal{H}(t - t_2)] + \frac{U(\tau, T_m, T_M)}{T_M - T_m}\ U(t, t_1, t_2) \quad (4.26)$$

The resulting expected demand (eq. (4.18)) is illustrated on Figure 4.12.

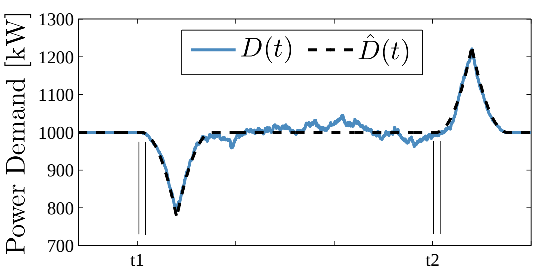


Figure 4.12: Evolution of the group's power demand following a transition from homogeneous to heterogeneous nature. Parameters: $P_n = 1kW, \lambda = 5s^{-1}, T_n = 200s, T_m = 20s, T_M = 380s$

On Figure 4.12, the plain blue line $D(t)$ is the exact demand (individual simulation of each load) and the dotted black line is its estimated counterpart $\hat{D}(t)$. As can be observed, the demand suddenly drops from time $t = t_1$ and then recovers its steady-state level of $1000kW = \lambda P_n T_n$. A symmetric behavior is observed from time $t = t_2$.

The observed demand drop and rise are somehow counter-intuitive as loads have the same expected parameters at all time. Two twin vertical lines are shown. The left-hand lines indicate the start of each transition period t_1 and t_2 while the right-hand ones show the time from which the expected demand begins to diverge from steady-state, T_m instants later. The following developments describe the evolution of the expected demand during both transitions. More details can be found in appendix B.

4.4.1 From homogeneous to heterogeneous group

Basically, the phenomenon at stake is simple. Initially, the energy content of the group is concentrated in the interval $[t - T_n, t]$. In the heterogeneous case, a single load has, on average, the same energy content than in the homogeneous case. However, at group-level, this energy needs to be spread on a longer time interval $[t - T_M, t]$. Therefore, the group-level power must decrease temporarily to *store* the energy on a longer time horizon.

Let's define time $x_1 = (t - t_1)$. In the first transition period, $0 \leq x_1 < T_M$, the dynamical behavior of the expected demand is derived below ($[a]_+$ conserves only the positive part of a).

$\forall x_1 \in [T_m, T_M]$

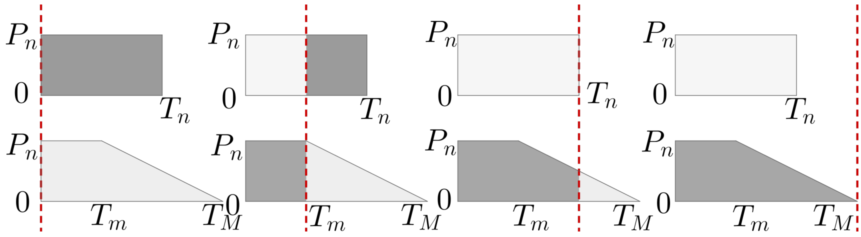
$$\hat{D}(t) = \lambda P\left([T - x_1]_+ + T_m + \frac{(T_m - x_1)(T_m + x_1 - 2T_M)}{2(T_M - T_m)}\right) \tag{4.27}$$

and $\hat{D}(t) = \lambda PT$ if $0 \leq x_1 < T_m$ or $T_M < x_1$

The observed variations of $\hat{D}(t)$ are illustrated on Figure 4.13. Four different charts are shown corresponding to different instants within the transition. The vertical dotted line represents the time, the top chart is the initial load consumption profile, and the bottom chart is the heterogeneous profile. Both profiles have the same total area (energy content). However, during the transition, the average energy content of a running load profile will vary. It is represented by the the dark area which sums the right-hand side (w.r.t. time t) of the homogeneous profile area (top) with the left-hand side of the heterogeneous profile area (bottom). The dark area area starts shrinking from time $x_1 = T_m$ and recovers from time $x_1 = T_n$. The energy content is constant in the interval $x_1 \in [0, T_m]$.

4.4.2 From Heterogeneous to homogeneous group

From time $0 \leq x_2 < T_M$, a similar dynamical behavior leads to a symmetric increase of the expected power. It is described in the following equation,



(a) $x_1 = 0$ (b) $x_1 = T_m$ (c) $x_1 = T_n$ (d) $x_1 = T_M$

Figure 4.13: Average energy content of the running loads during the transition from homogeneous to heterogeneous groups. Top: constant power (homogeneous). Bottom: expected power profile $\omega(\tau)$ (heterogeneous). The dark grey area is the expected demand of the group $\hat{D}(t)$ (normalized by λ) at each time.

where $x_2 = (t - t_2)$ and $\left(x_2|_{T_m}^{T_M}\right)$ is $\min(\max(x_2, T_m), T_M)$.

$$\hat{D}(t) = \lambda P\left(T + [T_m - x_2]_+ + \frac{T_M^2 - 2T_M\left(x_2|_{T_m}^{T_M}\right) + \left(x_2|_{T_m}^{T_M}\right)^2}{2(T_M - T_m)}\right) \tag{4.28}$$

4.5 Demand volatility

The objective of this section is to assess the influence of parameter and arrival randomness on the group's demand volatility. We want to specify the set of parameter values that ensures a sufficiently limited volatility of the demand around its expected value.

Indeed, our final goal is to control ECLs in order to provide a measurable response to frequency deviations. We want ECLs to respond autonomously to frequency changes providing accurate response. For instance, the group as a whole should be able to consume at a lower, predetermined level, with respect to its initial level in case frequency is below nominal (i.e., 50Hz).

The main limitation of autonomous control is that loads have no direct access to the group-level information. Only indirect information is available. A load could get an image of the group-level state if it can be derived from past local measurements or past control efforts of the load itself. However, the detectable information will likely be limited to the expected level of the group demand and states. All variations around this expected value would require much more direct information, in the form of measurement feedback, to be accessible to each load.

In control systems, the only way to counteract the effect of a random perturbation is through the use of output and/or state feedback. An adequate modeling of the noise influence on the system output/states is also required. In fact, the past and present realizations of a random perturbation form a deterministic process. Therefore, their effect on the system states are similar to the effect of control inputs that require measurement feedback.

In conclusion, autonomous control will not behave adequately in groups whose natural demand volatility is large compared to the amount of flexible capacity the group aims at providing. In other words, the relative volatility of the power demand $D(t)$ around its expected value should be small.

Let's define a volatility metric v_D that will be used to define the interval in which the group-level demand $D(t)$ stays most of the time. The volatility v_D is the ratio between the standard deviation $\sigma_D = \sqrt{\mathrm{Var}[D]}$ and the expected value (baseline) $D_0 = \mathbb{E}[D]$ of the demand distribution.

$$v_D = \frac{\sigma_D}{D_0} \tag{4.29}$$

We would like to assess the number of loads nl (the group size) and parameters that could guarantee that the group volatility is acceptable. To this end, we define the *volatility interval* $D_0(1 \pm 3v_D)$. The group's demand lies in this interval *most of the time* (see below). In order to guarantee good control performances, the group should be large enough such that its demand *in the uncontrolled case* varies most of the time in an interval $3v_D \leq 1 - 2\%$.

The number of ECLs nl encompasses both running and idled loads. Assuming that users start their appliance on average once every two days ($Ud = 0.5$) and that load arrivals occur at regular rate, we can derive the link between the arrival rate λ and the total number of loads in the group nl: $\lambda = nl(Ud/86400)$.

The volatility interval $1 \pm 3v_D$ is chosen according to the following reasoning. In the next chapters, we will exploit the relative demand change $x(t)$, expressed in *per unit* w.r.t the expected level D_0: $x(t) = D(t)/D_0 - 1$. Let's choose a number $\mathrm{n} \in \mathbb{R}$ that defines the probability $p^{in}(\mathrm{n})$ that $x(t) \in [-\mathrm{n}v_D, \mathrm{n}v_D]$. The frequency of *outside range* observations $N^{out}(\mathrm{n})$ corresponds to the inverse of the complementary probability $p^{out}(\mathrm{n}) = 1 - p^{in}(\mathrm{n})$.

$$N^{out}(\mathrm{n}) = \frac{1}{1 - p^{out}(\mathrm{n})} \tag{4.30}$$

Table 4.1: Number of outside range observations of a second-based event on a yearly basis (normal distribution).

n	$N^{out}(\mathrm{n})[y^{-1}]$	Description
1	100 mio	Most of the time.
2	1.5 mio	2-3 times a minute.
3	85000	10 times an hour.
4	2000	5-6 times a day.
5	180	Once every two days.
6	0.6	Least than once per year.

When normally distributed, second-based event would be observed a number of times $N^{out}(\mathrm{n})[y^{-1}]$ on a yearly basis. The chances of seeing the relative demand outside the range $[1 \pm \mathrm{n}v_D]$ are computed with the error function, and presented on table 4.1.

$$N^{out}(\mathrm{n}) = 31536000(1 - erf(\frac{\mathrm{n}}{\sqrt{2}}) \tag{4.31}$$

The choice of n and the performance metric $\mathrm{n}v_D \leq 1\%$ or $\mathrm{n}v_D \leq 2\%$ are complementary. In what follows, we will consider n = 3, which leads to observe out-of-range events on average once every 5-6 minutes. We will explore the set of load/group parameters that guarantee $3v_D \leq 1\%$, which is rather strict.

4.5.1 Queuing Theory for volatility estimate

Queuing theory is a branch of mathematics and statistics that study the dynamics of queues. Queues $A_r/T_s/N_s$ (see Fig. 4.14) may be any process in which some elements arrive randomly with arrival rate A_r and wait until they get *served*. The service time is random of distribution T_s. Clients get served by a defined number of servers N_s.

The number of loads running $N(t)$ in the group at time t can be described as a queuing process. A load is considered *in service* (i.e., has access to a *server*) when it consumes energy. Therefore, the number of servers N_s is infinite, as a load can start consuming energy right after it has arrived (i.e., no waiting time).

The group may be represented with specific kind of queues depending on the randomness nature of its parameters and/or load arrivals. Indeed, the inter-arrival intervals can be deterministic (D), exponentially distributed

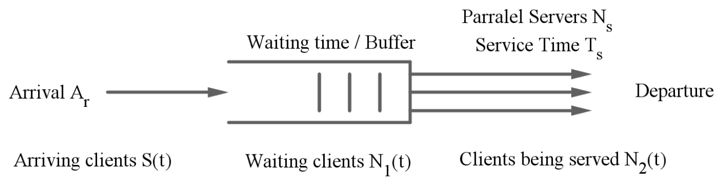


Figure 4.14: General $A_r/T_s/N_s$ Queuing Model.

(M) of follow a general type of distribution (G), The same notation applies for the service type. In particular, *Markovian* queues ($M/\cdot/\cdot$) are characterized by Poisson arrival processes. These are particularly important due to their memoryless property, as mentioned above. Looking at the number of loads running in a load groups, four types of queues are of interest.

1. $D/D/\infty$ queue. This queue is characterized by deterministic arrival process and service time. In such case, the number of running loads is constant.

2. $M/D/\infty$ queue. Variable starting process with deterministic service time. This queue represents an homogeneous group.

3. $D/G/\infty$ queue. Deterministic starting process with parameter distribution of general type.

4. $M/G/\infty$ queue. This queue represents realistic load groups with random starting process and random parameters.

Distribution of a sum of a random number of random variables

Queuing theory is helpful in studying the characteristic of the number of running loads $N(t)$. Some additional assumption are needed to characterize the randomness of the power rates. We want to characterize the expected value and variance (first and second moments) of the sum of these random variables.

The power demand of the group is the sum of a random number of random variables. The random number of elements in the sum is the number of running loads. The random variables represent the power rate at which each of these loads consumes. The demand of the group consist in a sum of random variables Z_i (power rate) associated to each one of the $i = 1..N(t)$

running loads.

$$D(t) = \sum_{i=1}^{N(t)} Z_i \tag{4.32}$$

Assuming the random variable Z_i's to be mutually independent and identically distributed, a powerful tool denoted moment-generating function (MGF) can be used to derive all moments of a sum of random variables. The moment-generating function $M_Z(t)$ of a variable Z is the expected value of an exponent that take as argument the product of time t with variable Z. It can be shown that it corresponds to an infinite sum of the *moments* μ_k of Z [65]. The n-th time derivative of the MGF evaluated at $t = 0$ corresponds to n-th moment μ_n. The zeroth moment μ_0 of Z is the integral of the distribution function and is always equal to one. The first and second moments are the mean and variance, respectively.

$$M_Z(t) = \mathbb{E}[\exp(tZ)] = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} \tag{4.33}$$

$$\mu_n = \frac{d^n}{dt^n} M_Z(t)\Big|_{t=0} \tag{4.34}$$

Here is where the MGF turns useful: the moment-generating function $M_{S_N}(t)$ of a sum S_N of N independent random variables $Z_i, \forall i = 1..N$ is the product of the N individual MGFs. As N is random with known probability to be equal to any positive integer $k \in \mathbb{Z}^+$, the MGF of the sum is a weighted sum of products of MGFs.

$$M_{S_N}(t) = \sum_{k \in \mathbb{Z}^+} Pr\{N = k\} \prod_{i=1}^{k} M_{Z_i}(t) \tag{4.35}$$

The first moment (expected value) of S_N is easily derived from the moments relative to Z_i.

$$\mathbb{E}[S_N] = \frac{dM_{S_N}}{dt}(0) \tag{4.36}$$

$$= \sum_{k \in \mathbb{Z}^+} Pr\{N = k\} \sum_{i=1}^{k} \frac{dM_{Z_i}}{dt}(0) \prod_{j \in [1,k], j \neq i} M_{Z_j}(0) \tag{4.37}$$

$$= \sum_{k \in \mathbb{Z}^+} Pr\{N = k\} \sum_{i=1}^{k} \mu_{1,i} \tag{4.38}$$

where, $\mu_{1,i}$ is the first moment of the i-th variable. For i.i.d. Z_i's of mean $\mathbb{E}[Z]$, the first moment of the sum is given below.

$$\mathbb{E}[S_N] = \mathbb{E}[Z] \sum_{k \in \mathbb{Z}^+} kPr\{N = k\} = \mathbb{E}[Z]\mathbb{E}[N] \tag{4.39}$$

The second moment (variance) $\text{Var}[S_N]$ is given below.

$$\text{Var}[S_N] = \frac{d^2 M_{S_N}}{dt^2}(0) \tag{4.40}$$

$$= \sum_{k \in \mathbb{Z}^+} Pr\{N = k\} \Big[\sum_{i=1}^{k} \frac{d^2 M_{Z_i}}{dt^2}(0) + \sum_{i,j \in [1,k], j \neq i} \frac{dM_{Z_i}}{dt}(0) \frac{dM_{Z_j}}{dt}(0) \Big] \tag{4.41}$$

$$= \sum_{k \in \mathbb{Z}^+} Pr\{N = k\} \Big[\sum_{i=1}^{k} \mu_{2,i} + \sum_{i,j \in [1,k], j \neq i} \mu_{1,i}\mu_{1,j} \Big] \tag{4.42}$$

$$= \sum_{k \in \mathbb{Z}^+} Pr\{N = k\} \Big[\sum_{i=1}^{k} \mu_{2,i} + (\sum_{i \in [1,k]} \mu_{1,i})^2 - \sum_{i \in [1,k]} \mu_{1,i}^2 \Big] \tag{4.43}$$

where, $\mu_{2,i}$ is the first moment of the i-th variable.

In case variables are identically distributed with variance σ_Z^2, the second moment of the sum is as follows.

$$\text{Var}[S_N] = \text{Var}[Z]\mathbb{E}[N] + \mathbb{E}[Z]^2 \big(\mathbb{E}[N^2] - \mathbb{E}[N] \big) \tag{4.44}$$

4.5.2 $M/D/\infty$: homogeneous group with random arrivals

Following equations (4.15) and (4.14), the relative volatility in a group with random arrivals but deterministic parameters is found below.

$$v_D = \frac{1}{\sqrt{\lambda T_n}} \tag{4.45}$$

These results are illustrated on figure 4.15. The natural run time of the group has a major influence on v_D. On the other hand, the required number of loads in the group nl depends largely on the average number use per day Ud. On figure 4.15, we have considered that loads are used on average once every two days $Ud = 1/2$.

In case loads run time is 2h, the group should count about a million loads for its demand volatility to be acceptable (i.e., $3v_D \leq 1\%$).

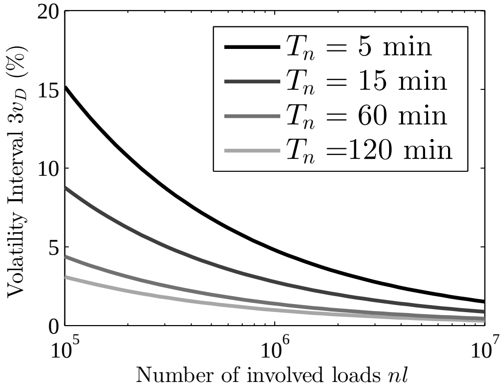


Figure 4.15: Half amplitude of the volatility interval $3v_D$ as a function of the group size and the load's run time (homogeneous group, random starts, $Ud = \frac{1}{2}$).

4.5.3 $D/D/\infty$ with Variable Power

We consider a group with deterministic arrivals for which λ loads arrive every second and run during a fixed time T_n. The power rate of these loads are randomly distributed. The group's demand group consists in a sum of $N = \lambda T_n$ random power rate P_i with mean P_n and variance σ_P^2. From equation (4.39) and (4.44), we deduce the mean and variance of the total demand.

$$D = \sum_{i=1}^{N} Z_i \quad \Rightarrow \quad \mathbb{E}[D] = D_0 = \lambda T_n P_n \quad \& \quad \text{Var}[D] = \sigma_D^2 = \sigma_P^2 \lambda T_n \tag{4.46}$$

The influence of different parameters on the group demand volatility is illustrated on figure 4.16. In all of the four illustrated case, the rate $\lambda = 1$ is the same. The loads have a run time equal to $1/4h$ (top charts) or $2h$ (bottom charts). The mean value of the power rate is normalized to 1. For convenience, we denote by $\sigma_P^r = \sigma_P / P_n$ the relative standard error of the random power rates. This latter is set respectively to 30% and 50% on the right and left-hand side charts (i.e., $\Delta P = 50\%$ and $\Delta P = 86\%$ respectively). The shaded areas represent $1 \pm 3v_D$ volatility interval. The half amplitude of this interval is a function of both λT_n and σ_P^r.

$$3v_D = 3\frac{\sigma_D}{D_0} = 3\frac{\sigma_P^r}{\sqrt{\lambda T_n}} = 3\frac{\sigma_P^r}{\sqrt{\frac{7800nl}{2\times 86400}}} \tag{4.47}$$

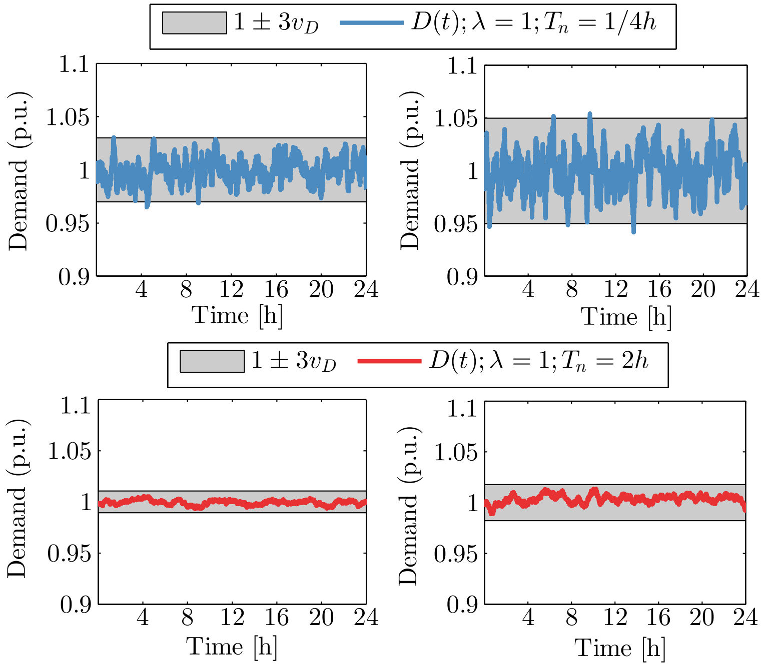


Figure 4.16: Illustration of demand volatility with random power rate (uniform distribution). Demand is shown in per unit. Left charts: $P \in 1 \pm 50\%$. Right charts: $P \in 1 \pm 86\%$.

On figure 4.17, we represent the relative volatility evolution with the number of involved nl loads. For basically any realistic value of the power relative standard error, a group counting about one million loads would experience a demand volatility interval below 1%.

4.5.4 $D/G/\infty$: Variable Run Time

In a group where only run time are variables, the group's dynamics can be represented by a $D/G/\infty$ queue. The number of loads running at time t is random, not because loads arrive at random rate, but because they stop after i.i.d. random run times $T \in [T_m, T_M]$, with known distribution Ω_T. Let's define an infinitesimal load arrival λdt to any infinitesimal interval $[t, t + dt]$. Each of the infinitesimal load that has arrived $\tau \in [0, T_M]$ instants before current time t has a probability $\Omega_{T \leq \tau}$ to be sill running. Consequently, the static distribution of the number of running loads N can be obtained by associating a Bernoulli trial of success rate $\mathrm{p}(\tau) = Pr\{T \geq \tau\}$

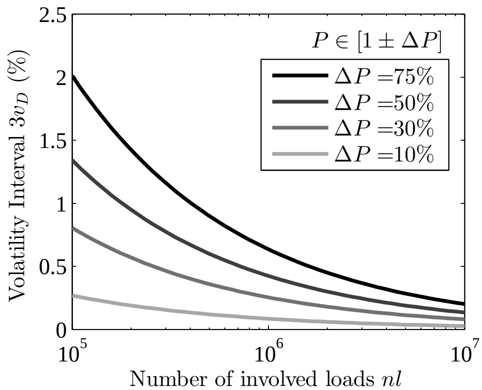


Figure 4.17: Half amplitude of the volatility interval $3v_D$ as a function of the relative standard error of random power P and the group size (Variable power, constant arrivals, run time $T_n = 2h$, $Ud = \frac{1}{2}$).

to each the infinitesimal load. The success rate is equal to 1 in the interval $\tau \in [0, T_m]$ and zero if $\tau > T_M$. Considering uniformly distributed run times, $\mathrm{p}(\tau) = \frac{T_M - \tau}{T_M - T_m}$ for all $\tau \in [T_m, T_M]$. This is illustrated on figure 4.18. The expected value and variance of N are obtained by integration of the infinitesimal expected value $dE(\tau) = \mathrm{p}(\tau)\lambda d\tau$ and variance $dV(\tau) = \mathrm{p}(\tau)(1 - \mathrm{p}(\tau))\lambda d\tau$, respectively.

$$\mathbb{E}[N] = \int_0^{T_M} dE(\tau) = \lambda T_n \tag{4.48}$$

$$= \lambda T_m + \lambda \int_{T_m}^{T_M} \frac{T_M - \tau}{T_M - T_m} d\tau = \lambda \frac{T_M + T_m}{2} \tag{4.49}$$

$$\mathrm{Var}[N] = \int_0^{T_M} dV(\tau) = \lambda \int_{T_m}^{T_M} \frac{T_M - \tau}{T_M - T_m} \frac{\tau - T_m}{T_M - T_m} d\tau \tag{4.50}$$

$$= \lambda (T_M - T_m) \int_0^1 z(1 - z) dz = \frac{\lambda (T_M - T_m)}{6} \tag{4.51}$$

Which gives,

$$D_0 = \lambda P_n T_n \quad \& \quad \sigma_D^2 = P_n^2 \frac{\lambda (T_M - T_m)}{6} \tag{4.52}$$

The relative volatility v_D of demand around its expected value evolves as a function of the relative standard error of the run time σ_T^r, and inversely with the square root of the number of running loads. It is therefore strongly

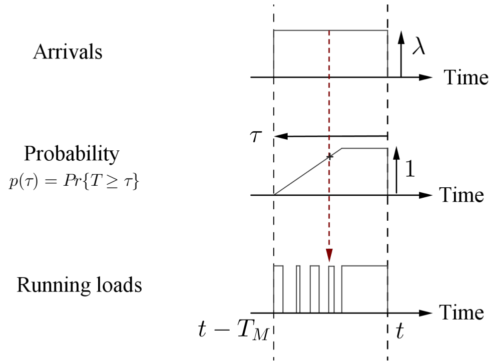


Figure 4.18: Number of running loads with random run times. The constant mass distribution of arrivals (top) passes through a Bernoulli trial filter of different success rates $p(\tau)$ (mid) and gives a random realization (bottom).

influenced by the average run time of the loads, as illustrated on figure 4.19.

$$v_D = \frac{\sqrt{\sigma_T^r}}{3^{1/4}\sqrt{\lambda T_n}}$$

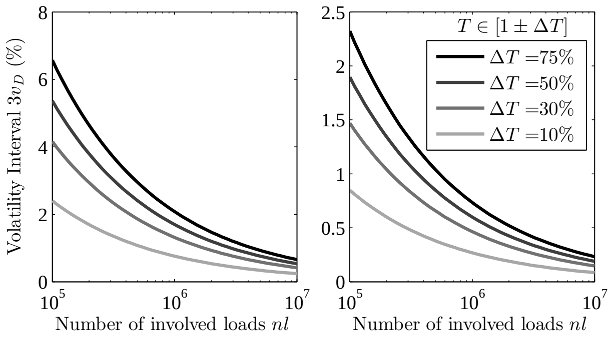


Figure 4.19: Half amplitude of the relative demand volatility as a function the number of loads nl and the standard error and mean value of the random load run time. Left:$T_n = 1/4h$. Right: $T_n = 2h$.

4.5.5 $M/G/\infty$: Variable Parameters and Starting Process

In realistic ECL groups, both arrivals and parameters are random. As shown in [110], the number of customers $N(t)$ being served in a $M/G/\infty$ queue with Poisson arrivals of parameter λ is distributed as follows.

$$Pr\{N(t) = m\} = \frac{\big(\lambda \int_0^t [1 - H(z)]\, dz\big)^m \; exp\big(- \lambda \int_0^t [1 - H(z)]\, dz\big)}{m!} \tag{4.53}$$

where, $H(z) = Pr\{\text{service time} \leq z\}$.

In steady-state, and considering an average service time (=run time) T_n, N is Poisson distributed.

$$\lim_{t\to\infty} Pr\{N(t) = m\} = \frac{(\lambda T_n)^m \; exp\big(- \lambda T_n\big)}{m!} \tag{4.54}$$

The moment-generating function (MGF) $M_{S_N}(t)$ of a sum $S_N = Z_1 + Z_2 + ... + Z_N$ of a sequence of N independent random variable is the product of the N MGFs.

$$M_{S_N}(t) = \sum_{i=0}^{\infty} Pr\{N = i\} \prod_{k=1}^{i} M_{Z_i}(t) \tag{4.55}$$

For such group of ECL, equation (4.35) gives the following.

$$M_D(t) = \sum_{i=0}^{\infty} \frac{(\lambda T_n)^i \; exp\big(- \lambda T_n\big)}{i!} \prod_{k=1}^{i} M_{Z_i}(t) \tag{4.56}$$

The first and second moments are obtained from equation (4.39) and (4.44), supposing that load parameters are iid and that the average power is P_n (i.e., $\mathbb{E}[Z] = P_n$). The demand corresponds to a compound Poisson process [12].

$$\mathbb{E}[D] \quad = \mathbb{E}[Z]\mathbb{E}[N] = \lambda P_n T_n \tag{4.57}$$

$$\text{Var}[D] = \lambda T_n(\sigma_P^2 + P_n^2) \tag{4.58}$$

The relative volatility of demand v_D is defined below.

$$v_D = \frac{\sqrt{(\sigma_P^r)^2 + 1}}{\sqrt{\lambda T_n}} \tag{4.59}$$

4.5.6 Comparing Volatility

We consider an ECL group with Poisson arrival process of parameter λ populated by loads of power $P \in [P_m, P_M]$ and run time $T \in [T_m, T_M]$ distributed with average values P_n and T_n and variance σ_P and σ_T respectively. In order to compare the relative importance of parameter variability, let's define a unique parameter σ as the relative variance of all load parameters.

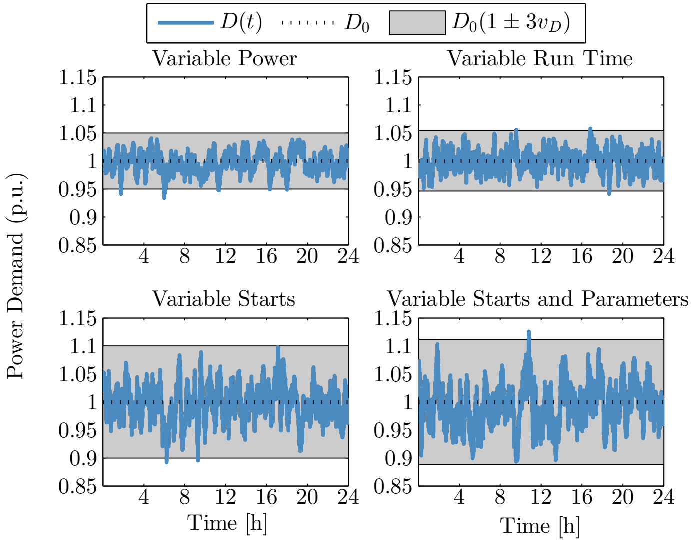
$$\sigma = \frac{\sigma_T}{T_n} = \frac{\sigma_P}{P_n} \tag{4.60}$$

We would like to compare the impact of the parameter distribution on the group's demand. Comparison of group-level relative volatility in function of different parameter randomness is presented on Table 4.2.

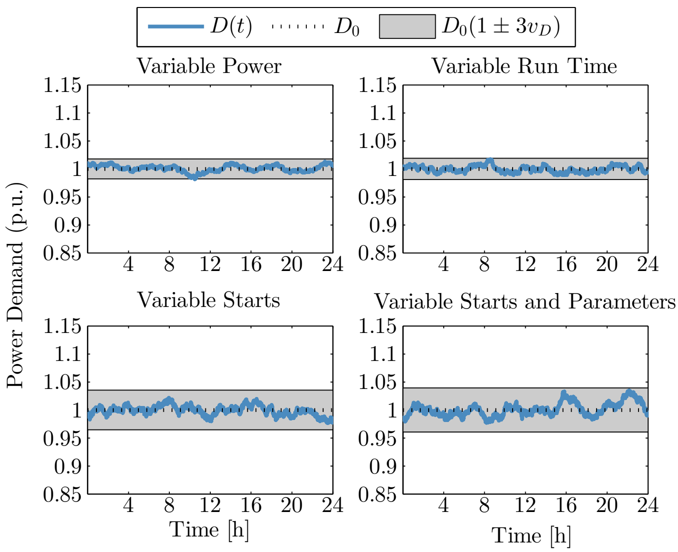
Table 4.2: Group-level consequences of different parameter randomness.

Random Variable	Power	Run Time	Arrivals	Altogether
Relative volatility v_D	$\frac{\sigma}{\sqrt{\lambda T_n}}$	$\frac{\sqrt{\sigma}}{3^{1/4}\sqrt{\lambda T_n}}$	$\frac{1}{\sqrt{\lambda T_n}}$	$\frac{\sqrt{\sigma^2+1}}{\sqrt{\lambda T_n}}$

Some illustrative examples are shown on figure 4.20 and 4.21. Both figures are relative to group where parameters are uniformly distributed with standard deviation $\sigma = 30\%$. Parameters may vary $\pm 50\%$ around their average value. The figure 4.20 shows a group with high starting rate $\lambda = 4$ and figure 4.21 with starting rate $\lambda = 0.1$. The smoothing effect of longer average run time is clearly visible from comparing the top and bottom charts within each figure.

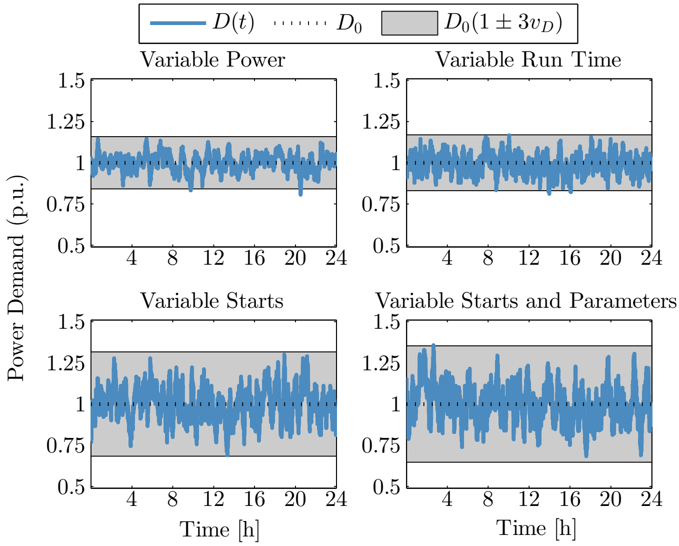


(a) $T_n = 1/4$h.

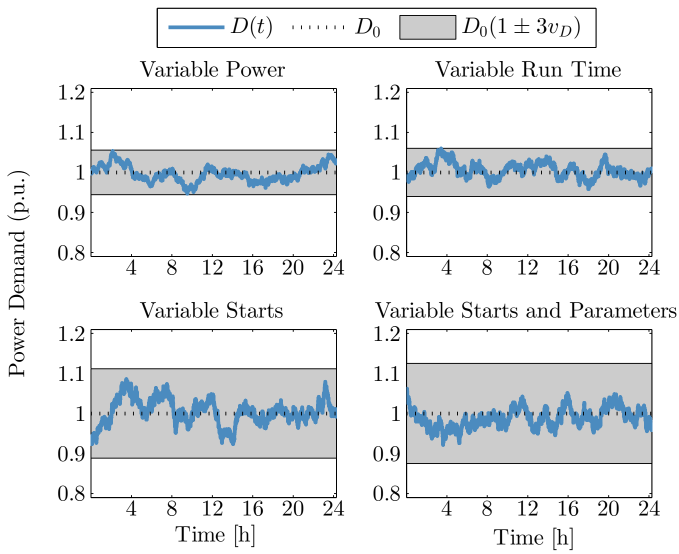


(b) $T_n = 2$h.

Figure 4.20: Illustration of relative demand volatility ($\lambda = 4s^{-1}, \sigma = 30\%$). Plain lines represent the demand of groups in which power, run time and/or arrivals are of random nature. Dotted lines are the expected level of demand. The shaded area is the volatility interval $1 \pm 3v_D$. The top chart and bottom chart represent respectively ECLs with 1/4h and 2h average run time.



(a) $T_n = 1/4$h.



(b) $T_n = 2$h.

Figure 4.21: Illustration of relative demand volatility ($\lambda = 0.1s^{-1}, \sigma = 30\%$). Plain lines represent the demand of groups in which power, run time and/or arrivals are of random nature. Dotted lines are the expected level of demand. The shaded area is the volatility interval $1 \pm 3v_D$. The top chart an bottom chart represent respectively ECLs with 1/4h and 2h average run time.

4.6 Chapter conclusion

In this chapter, we discussed the behavior of a large group of Energy Constrained Loads at rest (uncontrolled loads). It consisted in an important premise to the discussions of the following chapters, where we explore the dynamics of ECL group when its loads consumption is remotely controlled. In theory, the ECLs could be controlled in order to deliver flexibility to the electrical network they are connected to. Yet, the power demand of an ECL group is variable. From the system operator side, it is crucial that the demand variability is small compared to the flexible capacity offered by the group. Otherwise, the system operator would not be able to distinguish random variations from the actual delivered flexibility volumes.

The demand variability is a consequence of the random load arrivals as well as the random nature of their parameters (run time and power rate). We have seen that the randomness of both arrivals and run time have a much larger impact on the group demand variability than the power rate randomness. Furthermore, the larger the group, the smaller the relative variations of the group demand around its expected value (baseline). This leads to conclude that a group of sufficient size will insure variability to be small enough.

One could argue that flexible loads are capable of flattening their own demand in order to limit this variability. However, this would induce two negative impacts. Firstly, the power demand of the group has to be measured precisely. This requires the use of a dedicated infrastructure, measuring in real-time the (non)consumption of all involved loads, even those that are idled. Though technically possible, such measurements are costly in both financial and environmental terms (e.g., extra consumption of energy, infrastructure). Developing strategies that do not require such measurements is an advantage. Secondly, loads would need to assign part of their flexibility for this specific purpose. This would be detrimental to the potential service offered to the grid operator. It's seems therefore beneficial to address groups whose demand variability is small enough by nature.

In this chapter, we found that groups of ECL should be very large in order for their *volatility* (defined above) to reach the 1% threshold (w.r.t. the group's baseline). A group that counts a *million ECLs* (both running or idled) whose run times are *at least a quarter hour* and that are started on average once every two days by their user have sufficiently low variability. In conclusion, the rest of our discussions will consider massive implementation programs.

Chapter 5

Autonomous ECL power control providing FCR

Chapter summary

Resilience, the ability to cope with change. Across all domains, from physics to biology, from economy to psychology, it constitutes one of most desirable feature within the object of study. It allows systems to reconfigure and stabilize after disturbances, it drove Humans up to the top of the food chain, it limits the impact of economical chocs and makes it possible to recover from trauma. Small electric appliances developing an intrinsic ability to cope with change would contribute to the system's self-stabilizing nature. Resilience does not condone failures. Autonomously, detached from all infrastructures but the network itself, shall load control lead to resilience.

Our objective in this chapter is to develop aggregate models accurately representing the power demand dynamics of an ECL group whose loads are externally controlled. The idea is to exploit simple mathematical structures instead of detailed simulation of each load in the group. Such models are however unable to capture the volatility of the aggregate demand around its baseline and are therefore restricted to groups of sufficient size. In this work, ECLs are controlled for delivering frequency control, and more precisely Frequency Containment Reserves.

5.1 Autonomous FCR: local and group-level impacts.

In this chapter, we discuss the most simple control schemes that would allow ECLs to provide FCR in an *autonomous* framework. The objective of FCR is to provide a proportional and precise response to the system frequency deviation from nominal. This is known as *droop-based* control, where a certain proportional factor K_{FCR} (i.e. the droop, expressed in MW/Hz) drives the amount of flexible capacity that must be deployed to counteract a system frequency deviation $(f(t) - f_n)$. The deployed FCR is positive if it requires an increase of generation output, that is during under-frequency event (see figure 5.1).

$$FCR(t) = -K_{FCR}(f(t) - f_n) \tag{5.1}$$

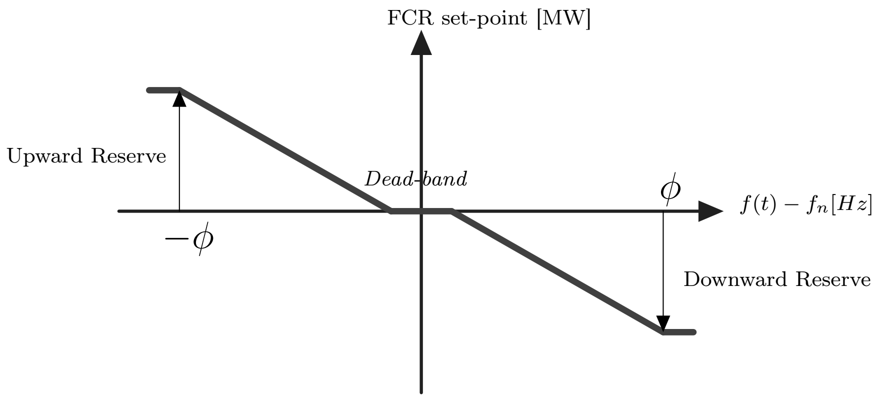


Figure 5.1: The power-frequency curve defining the FCR set-point.

The total capacity provided by all FCR participants is limited. In order to take this into account, a bounded and normalized version of the frequency deviation is used as reference signal $r(t) \in [-1, 1]$. A normalization frequency deviation ϕ is exploited and corresponds to the maximum allowed steady-state frequency deviation in the system being modeled (i.e. $\phi = 200mHz$ in Continental Europe [52]).

$$r(t) = \left. \frac{f(t) - f_n}{\phi} \right|_{-1}^{1} \tag{5.2}$$

The upward (resp. downward) FCR capacity is $K_{FCR}\phi$ (resp. -$K_{FCR}\phi$) and is fully deployed if $r(t) = -1$ (resp. $= 1$). Upward reserves refer to either increased generation output or, symmetrically, decreased power demand. A group of loads providing FCR must adapt its demand $D(t)$ proportionally to this reference frequency signal. We introduce the droop K_D $[Hz^{-1}]$ that is directly dependent on the total capacity that the group can deploy ($= \pm K_D D_0$).

$$D(t) \to D_0(1 + K_D\, r(t)) \tag{5.3}$$

Throughout this chapter, we suppose that the group provides only asymmetric upward FCR. That is, the demand can only decrease in the short-run. This is mainly done for pedagogical reasons, that we justify below. We have $0 \leq D(t) \leq D_0$ and $0 \leq K_D \leq 1$. A reference signal $r^{up}(t)$, restricted to negative frequency deviations, is defined.

$$r^{up}(t) = \left. \frac{f(t) - f_n}{\phi} \right|_{-1}^{0} \tag{5.4}$$

The group demand is influenced by local control decisions taken autonomously by each one of the $N(t)$ running loads. Loads have no access to communication infrastructure. The information locally available to each running load $i \in [1, N(t)]$ is limited to the locally measured frequency $f(t)$ and to the load's own parameters. Exploiting solely such local information, each load must be able to take a decision that will influence the group-level demand and make it frequency-responsive. As the number of running loads is not known, the load cannot know the exact amount of flexibility that its group provides. It can however estimate the *relative* flexibility $\hat{x}(t)$ offered by its group. That is, the relative change of the group demand w.r.t. the baseline level D_0.

$$\hat{x}(t) = \frac{\hat{D}(t)}{D_0} - 1 \rightarrow K_D\, r^{up}(t) \tag{5.5}$$

The total capacity that is actually provided by the group ($= -K_D D_0$) will be time varying. Indeed, in practical cases, the baseline demand D_0 is not constant and is dependent on the starting rate $\lambda(t)$. We assume that the system operator will take this variability into account in its operational planning, and use loads only when they are naturally available. Furthermore, we assume that the variations of the baseline demand are slow enough such that they can be neglected in the short-run.

After local controllers have been defined, the provided flexibility $\hat{x}(t)$ must be evaluated. To this end, the development of aggregate models is required. Indeed, there exist several practical cases where a detailed simulation of individual loads is both useless and computationally expensive. An aggregation model should represent accurately the group-level power demand dynamics as a weighted sum of local-level load parameters, behavior and control laws. The accuracy of the proposed model is assessed at the end of this chapter.

A special attention must be paid on the influence of energy and user-related constraints on the amount of flexible capacity that can be provided by the group (i.e. the droop K_D). This will be exploited in the next chapters for simulation purpose and economic analysis.

5.2 The Controller Design Problem

The main challenge of local controller design is to map local decisions to global demand change. The load must guess which decision is the best in order for its group to adapt its demand to the reference signal $r(t)$.

5.2.1 Autonomous Control Policies

In order to influence the relative flexibility $\hat{x}(t)$, each load is equipped with a power controller. This local controller is designed to (1) condition local information and (2) trigger the load's reaction, if needed.

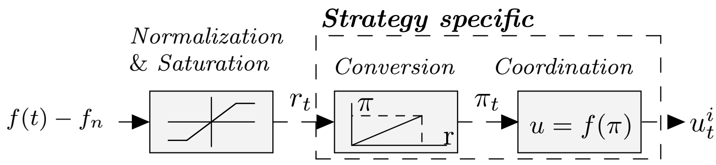


Figure 5.2: Control setup for a load i.

Three local ECL control policies are considered. They are described here below and illustrated on figure 5.3. These are based on the ECL definition presented in the introduction of chapter 4.

1. **Delay**. The starting instant of a load may be postponed. The largest applicable delay is $T_{dl} - T_n$ as the required energy must be consumed before the user deadline.

2. **Stop**. The load is either switched on ($P(t) = P_n$) or idled ($P(t) = 0$). The total idled time is limited to $T_{dl} - T_n$.

3. **Rate**. Load's power is set freely within a continuous interval $P(t) = P_n + b(t)P_n$. To respect the user deadline, one can impose a (restrictive) lower limit $P(t) \geq E_n/T_{dl}\, \forall t$.

For clarity, let's firstly insist on the distinction between the load arrival time and the load starting time. That is, a load arriving at time t and delayed by d instants actually starts at time $t + d$. A load that has arrived but has not yet started is said to be *idled*.

In this work, the local control decisions are defined as *incremental*. This means that loads will take a decision that will be applied during the following time-step (and will apply only to this time-step). A new decision occurs at each time-step.

As explained above we focus on loads providing upward FCR. Downward FCR is possible, but the comparison among the different control policies is less straightforward. For the sake of clarity, we found more useful to focus on the most simple considerations in order for the reader to acquire a fundamental understanding of the group behavior.

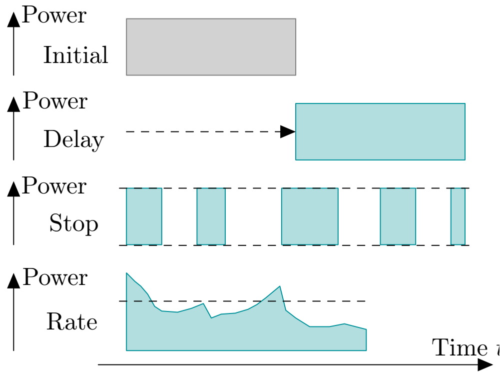


Figure 5.3: Control policies with energy conservation: (1) Delaying start, (2) Binary On/Off switching, (3) continuous power rate modulation.

A distinction between generic and selective policies

Depending on the policy at stake, the control can be either *selective* or *generic*. A generic load control policy applies the same decision to all loads that are in the same state (e.g., power rate, energy level, etc.). A selective control policy act on a portion of the loads that are in the same state. The Stop control policy is *selective*. Indeed, all running loads are, in principle, running at the next time step except if they decide otherwise. In order to select which of the running loads should be idled on the next time-step, the loads must exploit a random decision process. Each load carries out a Bernoulli trial that decides on it states on the next time-step (i.e., binary result 1 or 0). The success rate of these trials varies according to the measured reference signal.

On the other hand, both Delay and Rate policies can be either selective or generic. The Rate policy is selective if all loads can freely choose their set-point independently. It will be generic in case the power rate is defined in the same way for all running loads. The delay policy is selective if a portion of the arrived loads (i.e., that did not start yet) can decide to delay their start of one unit of time. It will be generic if all loads idled for the same amount of time take the same decisions. In the rest of this chapter, we define both rate and delay policies to be of generic type.

To define their control input (decision), loads will use the *coordination functions* that detailed in Table 5.1. Each coordination function exploits a specific trigger to generate an action: a discrete threshold, random number generation or a proportional controller. Decisions are performed at regular pace (e.g., every second) and maintained during one unit of time.

Table 5.1: Incremental (1 unit of time) decision description.

Policy	Delay	Stop	Rate
Decision impact	Delay start	Stop load	Change power
Decision type	Binary	Binary	Continuous
Decision variable	$u_t^{i,d}$	$u_t^{i,s}$	$u_t^{i,r}$
Decision trigger	Discrete Threshold	Random Number	Proportional

5.2.2 Energy rebound, remaining energy, stored energy

Before entering into further controller design discussions, we would like to introduce several important concepts. These emerge from the energy conservation nature of the proposed control framework.

When its loads are suddenly stopped/delayed/modulated, the group's demand will necessarily decrease temporarily as the load consumes less than expected. However, due to the energy constraint, the load will run for longer than initially scheduled. This results in additional amount of energy consumed after the load naturally stopping time. This is effect is known as the *energy rebound*.

Before this rebound occurs, the control allowed part of the scheduled energy consumption be shifted in time. It has therefore virtually stored energy *in time*. In consequence, the energy state of the group must necessarily have evolved in some ways. The energy state will recover its initial level after the rebound has occurred. It must must therefore be defined. In this work, we define two complementary energy states: The *remaining* energy and the *stored* energy of the group.

The remaining energy of the group E_r corresponds to the instantaneous energy level of its loads. At arrival, a load expects to consume an energy E_n, which is added to the group. After having started since $\Delta T \leq T_n$ instants, an energy $E_n - P_n \Delta T$ remains to be spent before the load can stop. In a non-controlled homogeneous group with constant arrivals λ, the sum of the remaining energy of each running load defines $E_{r,0}$, the remaining energy of the group at rest.

$$
\begin{aligned}
E_{r,0} &= \lambda \int_0^{T_n} E_n - P_n \tau \, d\tau = \lambda (E_n T_n - \frac{P_n T_n^2}{2}) = \frac{\lambda P_n T_n^2}{2} \\
&= \frac{1}{2} \lambda E_n T_n = \frac{1}{2} D_0 T_n
\end{aligned} \tag{5.6}
$$

The remaining energy $E_r(t)$ of the group will vary when loads are controlled. The remaining energy of a controlled load i arrived at time $t = t_{a,i}$ is $E_n - \int_{t_{a,i}}^{t} P_i(\tau) d\tau$. The difference between the group remaining energy with its level at rest is the stored energy $E_s(t)$.

$$
E_r(t) = \left(N(t) E_n - \sum_{i=1}^{N(t)} \int_{t_{a,i}}^{t} P_i(\tau) d\tau \right) \tag{5.7}
$$

$$
E_s(t) = E_r(t) - E_{r,0} \tag{5.8}
$$

The time derivative of the stored energy is the difference between the baseline power D_0 and the actual aggregate demand $D(t)$.

$$
\frac{dE_r(t)}{dt} = \frac{dE_s(t)}{dt} = \lambda E_n - D(t) = D_0 - D(t) \tag{5.9}
$$

This will be proven and exploited in the last chapter of this work, but simple reasoning can explain this nice property. At every time t, an infinitesimal load λdt arrives bringing an energy $\lambda E_n dt$ to the group. At the same time, all running loads consume together an energy $D(t)dt$, that has *eaten* up part of their remaining energy. We therefore have $dE_r(t) = \lambda E_n dt - D(t)dt$, which leads to the above result, as by definition $D_0 = \lambda E_n$.

5.2.3 Experimental framework for controller design.

For intuitive explanation of the controller design, we will conduct, for each of the three control policies, two sets of virtual experiments. Firstly, we observe the impact on the group demand when one single load is controlled according to the policy at stake. The load is either (1) delayed by 1 unit of time, (2) stopped during 1 unit of time or (3) modulated to minimal power during 1 unit of time. In parallel, the same action is imposed to this same load until the user time deadline becomes binding. In the second set of experiments, we apply the control actions to all running loads.

In all of these experiments, we consider a group of ECL populated by loads having the following parameters: $T_n = 8s$, $P_n = 1kW$, $E_n = 8kJ$, $T_{dl} = 12s$. Due to the user deadline, a load may be delayed at most by $T_{dl} - T_n = 4s$. In the rate policy, the user deadline defines a minimal level for the instantaneous power $P(t) \geq P_n(T_{dl} - T_n)/Tn = 2/3P_n$.

Before each experiment takes place, the group is in steady-state with a demand equal to $D_0 = 8kW$. Indeed, one load arrives at each discrete time-step of length $\Delta t = 1$, i.e., $\lambda = 1[s^{-1}]$. The experiments are conducted on the unique load arriving at $t = 4$.

5.3 Control Policy 1: Delaying load starts.

The control policy studied in this section acts on the instant at which loads start. As will be shown below, the main limitation is that control actions can be exercised on a very limited amount of loads. Indeed, one must wait for a load to effectively request to start in order to delay this same start. This implies that the provided flexibility cannot vary instantaneously, in other words, that its ramp rate is bounded by the load arrival rate.

5.3.1 Observation 1: Delaying one load start.

The unique load arriving at $t = 4$ is delayed either for (1) one unit of time (i.e. 1s) or (2) $T_{dl} - T_n = 4s$. It then runs as it was supposed to, at constant power $P_n = 1kW$. The impact on the aggregate demand of the group is shown on figure 5.4.

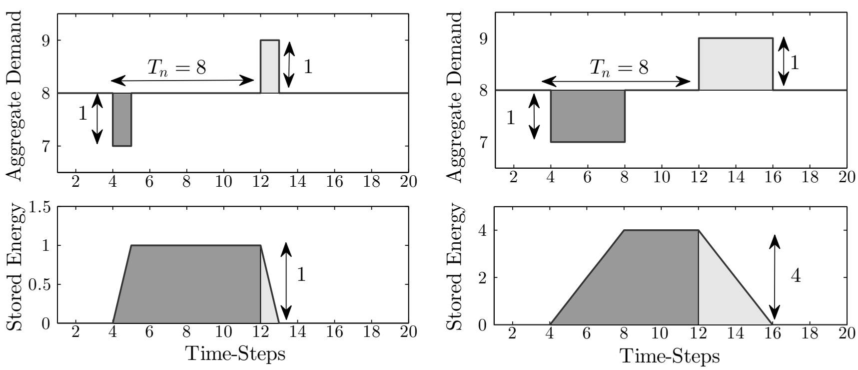


Figure 5.4: Single load impact: a delay is imposed before start (Delay Policy) during one unit of time (left-hand charts) and during the largest acceptable delay $T_{dl} - T_n = 4$ (right-hand charts).

The delayed load suddenly *disappears* from the group leading the aggregate demand to decrease to $D_0 - 1kW = 7kW$. If the delay lasts for 1 unit of time, the load starts immediately after and the aggregate demand returns its steady-state level. However, the load will need to run for one extra unit of time in order to consume the required energy E_n. As a consequence the aggregate demand rises up to $9kW$ precisely T_n instants after the delay decision was taken. In the second case, the same behavior is observed but both the demand decrease and the associated rebound last for $4s$.

In general, imposing a delay of length d to a load arrived at $t = t_1$ has the following impact on the arrival process, initially constant $S(t) = \lambda$.

$$\tilde{S}(t) = \lambda + \delta(t - t_1 + d) - \delta(t - t_1) \tag{5.10}$$

The impact on the aggregate demand can be expressed using the $T{-}wide$ time window function $U(t_1, T) = \mathcal{H}(t - t_1) - \mathcal{H}(t - t_1 - T)$, with $\mathcal{H}$ the Heaviside function.

$$D(t) = \lambda P_n \int_0^{T_n} \tilde{S}(t - \tau) d\tau = D_0 + P_n \left[U(t_1 + d, T_n) - U(t_1, T_n) \right] \tag{5.11}$$

Some energy was stored in the process and released back afterwards (rebound). During the transition phase $t \in [4, 4 + T_n + 1]$, and compared to a steady-state situation, the running loads have consumed temporarily less than the energy they were supposed to consume. Indeed, from its actual start $t = 4$ until its scheduled stopping time $t = 4 + T_n$, the delayed should have run more than it has. Its energy level should be closer to the target E_n. This phenomenon is shown on the bottom charts of figure 5.4. The group virtually stores energy when a load is delayed.

5.3.2 Observation 2: Delaying all arriving loads

We know look at the consequences of delaying all arriving loads. All loads arriving from time $t = 4s$ are delayed by the highest acceptable amount of time $T_{dl} - T_n$. This is illustrated on figure 5.5.

The demand of the group decreases proportionally with the number of loads that are delayed at each time-step. However, the energy rebound leads to an increase of the group demand precisely T_n instants after the initial delay decision. To model this behavior, let's denote by $w(t)$ the number of idled loads at time t (i.e., all loads that are waiting to start). At time $t = 4$, $\lambda \Delta t$ loads arrive and are delayed (i.e., $w(t = 4) = \lambda \Delta t$). At time $t = 5$, the same loads are delayed for an extra unit of time while another $\lambda \Delta t$ newly arrived loads are also delayed. This goes on forever, as all arriving loads are delayed until they reach their deadline. In such case, $w(t)$ increases until it reaches $w(t \geq (T_{dl} - T_n) + 4) = \lambda(T_{dl} - T_n)$.

$$w(t) = \lambda \Delta t \sum_{k=1}^{(T_{dl} - T_n)/\Delta t} \mathcal{H}(t - 3 + k\Delta t) \tag{5.12}$$

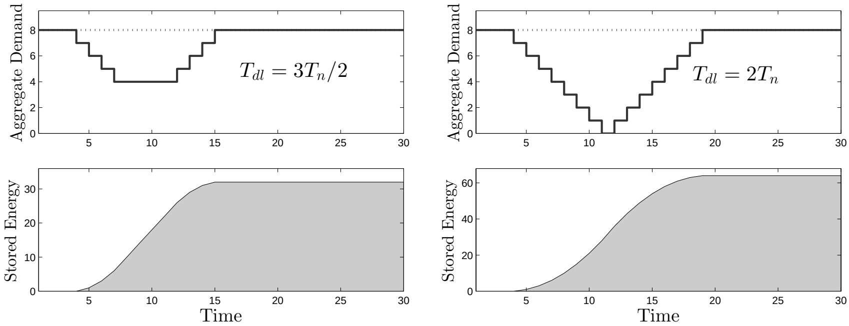


Figure 5.5: Highest acceptable time delay applied to all arriving loads from $t = 4s$. Parameters: $T_n = 8s$, $P_n = 1kW$, $\lambda = 1s^{-1}$. Left chart: $T_{dl} = 12s$. Right chart: $T_{dl} = 16s$.

The relative flexibility $x(t)$ offered by the group evolves as follows.

$$x(t) = \frac{w(t - T_n) - w(t)}{\lambda T_n} \tag{5.13}$$

Let's note that $x(t)$ is independent of the rate λ as all loads arriving at the same time follow the same behavior (i.e., $w(t)$ is proportional to λ). The amount of energy that can be stored by the group corresponds to the cumulative energy that was delayed by the idled loads. For readability, we prefer to express its normalized version (i.e., $\lambda = 1$, $P_n = 1$), expressed in $[s^2]$.

$$0 \leq \frac{E_s(t)}{\lambda P_n} \leq T_n(T_{dl} - T_n) \tag{5.14}$$

The above expression can be verified on the charts of figure 5.5 : the (normalized) stored energy rises up to $32[s^2]$ and $64[s^2]$ on the left and right charts respectively.

The number of loads $w(t)$ can be considered as a waiting queue. The maximum number of waiting loads is equal to $\lambda(T_{dl} - T_n)$. Therefore, the flexibility $x(t)$, and consequently the acceptable droop K_D are bounded.

$$x(t) > \max\left(1 - \frac{T_{dl}}{T_n}, -1\right) \quad \Rightarrow \quad K_D < \min\left(\frac{T_{dl}}{T_n} - 1, 1\right) \tag{5.15}$$

Furthermore, the ramp rate of the offered flexibility is bounded by the time derivative of $w(t)$. That is, the *additional* number of delayed loads at time t compared to $t-1$. As highlighted on figure 5.5, the relative demand of the group needs a time T_n to go from 1 ($D(t) = D_0$) to 0 when all arriving loads are systematically delayed. Indeed, we have,

$$\frac{dw(t)}{dt} < \lambda \quad \Rightarrow \quad \frac{dx(t)}{dt} > -\frac{1}{T_n} \tag{5.16}$$

According to ENTSO-E network code on frequency control and reserve [52], FCR power must be fully deployed in the 30s following an incident. This second element limits actually much more the FCR capacity ($= K_D$) that the group can offer.

$$K_D < \min\big(\frac{30s}{T_n}, \frac{T_{dl}}{T_n} - 1, 1\big) \tag{5.17}$$

As soon as the natural run time T_n is above 30 seconds (i.e., basically in all practical cases), the first limit is binding. This limit implies that only a tiny fraction of the baseline demand D_0 can be offered as FCR due to the ramp rate constraint. For instance, a group where loads run during 1 hour will only be able to control 1% of its baseline according to ENTSO-E's requirements. Obviously, the Delay control policy is not suited for FCR provision.

However, for slower services, the delay control policy could be useful. Therefore, it is still interesting to look for adequate controller design that would allow the group to autonomously adapt to a reference signal $r^{up}(t)$ with bounded time derivative.

The modified reference signal $\tilde{r}^{up}(t, T_n)$ takes this ramp rate limitation into account. It corresponds to a filtered version of the normalized frequency deviation, and depends on the load natural run time T_n.

$$\tilde{r}^{up}(t, T_n) = \min\big(\tilde{r}^{up}(t-1, T_n) - \frac{1}{T_n}; r^{up}(t, T_n)\big) \tag{5.18}$$

5.3.3 Local Controller Design

As shown above, a load i that decides to delay its starting time enters a *waiting queue* $w = \{0..w_l\}$ of maximal length $w_l = T_{dl} - T_n$ (i.e., user's constraint). Let's suppose that, before start, each arrived load observes the reference signal $r^{up}(t, T_n)$. The load can *target* a certain waiting time that can be derived from equation (5.13). Loads take binary (delay v.s. run) decisions on discrete time steps. These local decisions will be coordinated by a function of $\pi^d(t)$, a converted version of the reference signal $r^{up}(t, T_n)$,

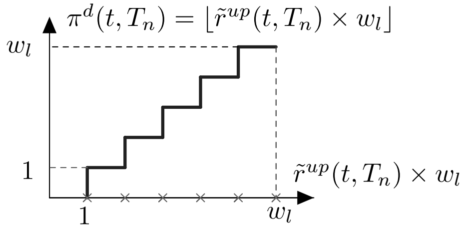


Figure 5.6: Conversion of $r^{up}(t, T_n)$ into $\pi^d(t, T_n)$ in the Delay policy.

where the superscript d stands for Delay policy. Neglecting the rebound (i.e., $w(t - T_n) = 0$), we have

$$\pi^d(t) = \frac{w(t|w(t-T_n)=0)}{\lambda} = -K_D T_n \tilde{r}^{up}(t) = -(T_{dl} - T_n)\tilde{r}^{up}(t) \qquad (5.19)$$

The coordination of each load is based on the integer-valued signal $\pi^d(t, T_n)$ that maps the continuous reference signal to target waiting times $[0, 1] \rightarrow \{0..w_l\}$. This is illustrated on figure 5.6.

The principle of local decisions is simple: a load i will wait an (extra) unit of time on the next time-step by comparing the target waiting time $\pi^d(t, T_n)$ to the time it has already spent being idled. The load may initially be in two states, either arriving at current time, or idled at previous time. The load decision is a binary value $u_i^d(t)$ that takes value 1 when the load is *delayed* at time t. Note that, as the group is homogeneous and frequency is the same everywhere in the system, the coordination function $\pi^d(t, T_n)$ is unique and shared by all loads (generic control policy).

$$u^{i,d}(t+1) = \begin{cases} 1, & \text{for arriving loads in case } \pi^d(t,T_n) \text{ is positive} \quad (5.20) \\ 1 & \text{for idled loads that waited less than } \pi^d(t,T_n) \quad (5.21) \\ 0, & \text{elsewhere} \quad (5.22) \end{cases}$$

5.3.4 Frequency response simulation

We simulate the reaction of the group during 3 hours. The reference signal $\tilde{r}^{up}(t)$ is the filtered version of the frequency deviations measured in Belgium on the 5th of January 2013, from midnight to 2 am (source : Elia System Operator). The signal is then set to zero for the remaining hour of the simulation in order to highlight the rebound effect. For illustration purpose, the initial frequency deviations are multiplied by a factor 3. Loads arrive on a regular basis and their parameters are deterministic (homogeneous group). The saturation limit ϕ is fixed to 200mHz. The results are shown on figure 5.7. The top chart shows the demand change target $K_D\tilde{r}^{up}(t)$, where $K_D = \frac{T_{dl}}{T_n} - 1 = 50\%$, and the actual demand change $x(t)$ (called *aggregate* response). On the bottom chart, the number of running loads $N(t)$ is compared to its steady-state level $N_0 = \lambda T_n$. The running loads increase is a consequence of past delay decisions combined with the energy constraint (i.e., $w(t - T_n)$). A similar simulation is shown on figure 5.8 for loads running twice as long. One should note the ramp rate change and the rebound occurring later.

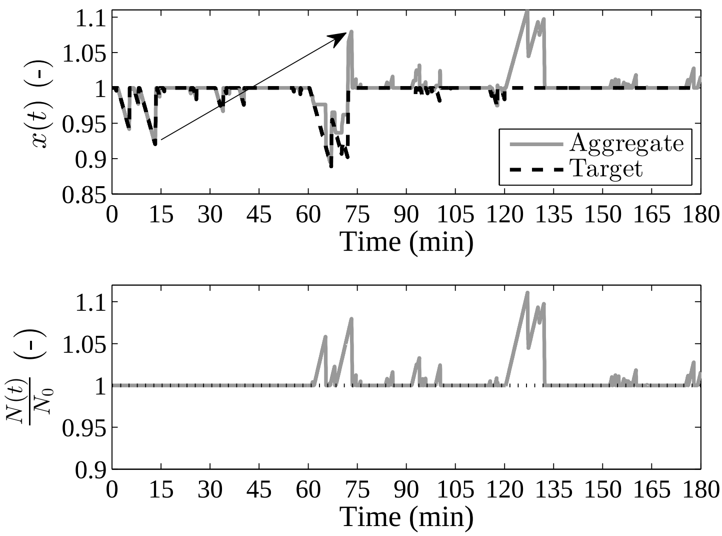


Figure 5.7: Group response to (filtered) frequency deviations in the Delay control policy. The arrow highlights the rebound impact of decisions taken T_n instant earlier, consecutive to the increase of the number of running loads (bottom chart). Parameters: $T_n = 1h$, $w_l = T_{dl} - T_n = 30min$, $\lambda = 1s^{-1}$, $P_n = 1kWh$.

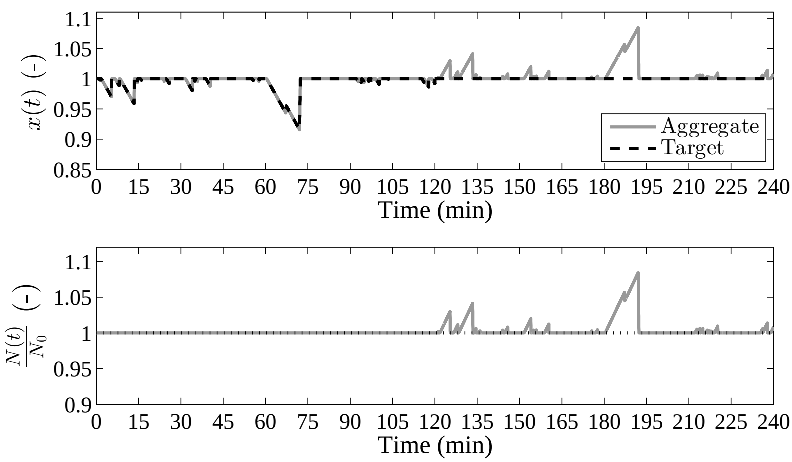


Figure 5.8: Group response to (filtered) frequency deviation in the Delay control policy. Parameters: $T_n = 2h$, $w_l = T_{dl} - T_n = 1h$, $\lambda = 1s^{-1}$, $P_n = 1kWh$. Loads run longer, which limits the relative ramp rate of the demand change.

5.4 Control Policy 2: binary On/Off switching.

The Stop policy is a more general version of the Delay policy. Indeed, it can act on any of the running loads.

5.4.1 Observation 1: Stopping one single load

We conduct an experiment in which the only load that has arrived at time $t = 4s$ is stopped during one unit of time (figure 5.9) and 4 units of time (figure 5.10). This load can be stopped at anytime during its run. The left-hand (resp. right-hand) charts of both figures show the consequences of stopping the load 2 seconds after its arrival (resp. 6 seconds), that is at time $t = 6$ (resp. $t = 10$).

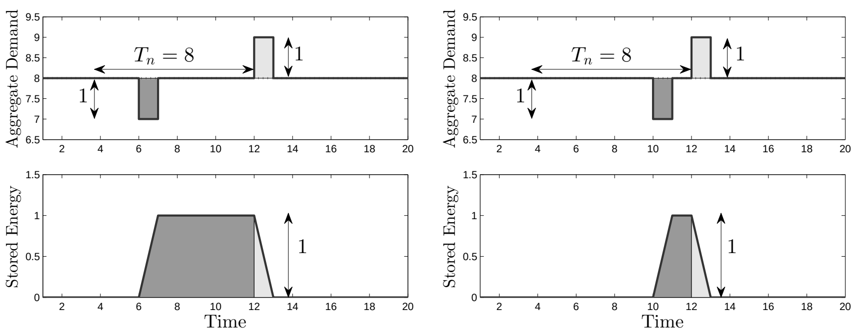


Figure 5.9: Single load impact (Stop Policy): one load is stopped in the middle of its run. The load is stopped during one second. Left-hand charts: stopped from 2 seconds after arrival. Right-hand charts: stopped from 6 seconds after arrival.

Stopping two loads of different energy state has a different impact on the group. This is illustrated on figure 5.10.

The right-hand chart of figure 5.10 corresponds to a case where the idled load, stopped during a delay $d = 4s$, has a remaining energy level $\Delta E = 2\Delta t P_n$ (i.e., the load has spent an energy equal to $6\Delta t P_n$ from its arrival). The maximum amount of energy that can be stored in the group gets thereby limited to this same value (=2). On the left-hand chart, the stored energy rises up to $dP_n = 4$, as the remaining energy of the idled load is not limiting in this case (=$6\Delta t P_n$).

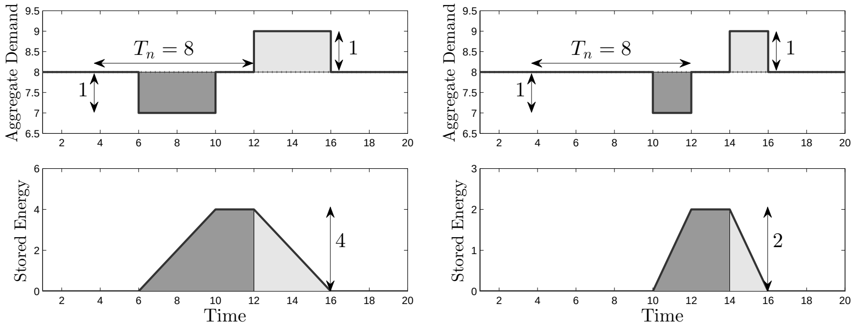


Figure 5.10: Single load impact (Stop Policy): one load is stopped in the middle of its run. The load is stopped during 4 seconds. Left-hand charts: stopped from 2 seconds after arrival. Right-hand charts: stopped from 6 seconds after arrival.

Consequently, we see that the maximum amount of energy that can be shifted by a single load being stopped during d instants is the minimum of three values (as $d \leq T_{dl} - T_n$).

$$E_s(t|d, \Delta T) < P_n \min(d, \Delta T, T_{dl} - T_n) \quad \text{valid for a single load} \tag{5.23}$$

In the above expression, $\Delta T = \Delta E / P_n$ is the time that remains to a load running at natural power before it has consumed the required energy E_n.

Let's suppose a load is stopped at time $t = t_1$ during a time d with a remaining time $\Delta T \in [0, T_n]$. Delaying a load of run time ΔT from $t = t_1$ during a time d would have the same impact on the group. We can therefore apply the same reasoning than in the Delay policy taking into account the remaining time ΔT of the load.

$$\tilde{S}(t|\Delta T) = \lambda - \delta(t - t_1) + \delta(t - t_1 + d) \tag{5.24}$$

$$D(t|\Delta T) = \int_0^{T_n} S(t-\tau)d\tau + \int_0^{\Delta T} \tilde{S}(t - \tau|\Delta T)d\tau \tag{5.25}$$

$$= D_0 - P_n U(t_1, \Delta T) + P_n U(t_1 + d + \Delta T, \Delta T) \tag{5.26}$$

Considering all possible ΔT, the average demand impact of stopping one load at time t_1 has an interesting profile, as can be seen on figure 5.11 (illustrated here for integer-valued $\Delta T = \{1, 2.., T_n\}$).

The immediate impact of a load stop is independent of its remaining time: the aggregate demand drops by $1kW$. The energy rebound is on

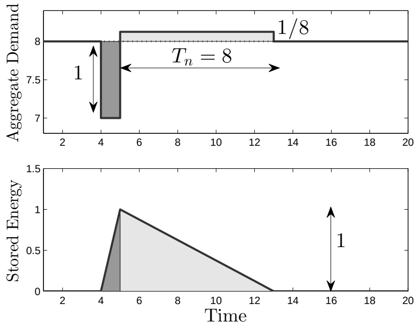


Figure 5.11: Average impact of a load stop on aggregate demand (Stop Policy).

average lower than what was observed in the rate policy, as it may occur on different time steps between $t_1 + 1$ and $t_1 + T_n$ depending on the load's remaining time. The average rebound magnitude is $P_n/T_n = 1/8kW$. It is much smaller than the initial demand impact of the load stop, and lasts on average T_n instants as energy is conserved.

5.4.2 Observation 2: Stopping all running loads.

The second set of observations focuses on the consequences of stopping all running loads until the user constraint becomes binding. Results are shown on figure 5.12.

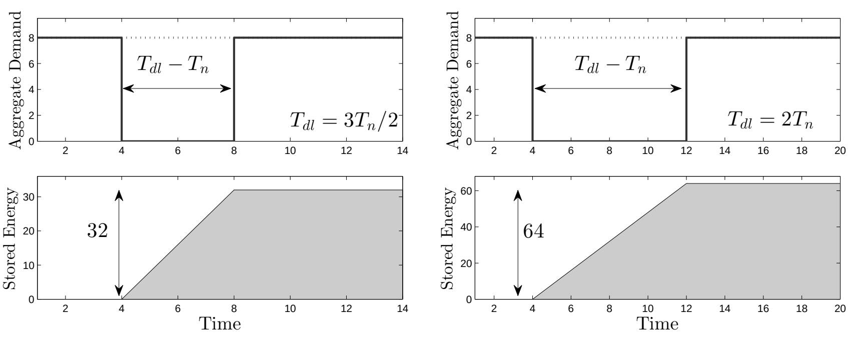


Figure 5.12: Stop of all running and arriving loads from $t = 4s$ until they individually reach their acceptable limit. Parameters: $T_n = 8s$, $P_n = 1kW$, $\lambda = 1s^{-1}$. Left chart: $T_{dl} = 12s$. Right chart: $T_{dl} = 16s$.

Each of the $N_0 = \lambda T_n$ running load can store an energy equal to $P_n(T_{dl} - T_n)$ before the loads reach their constraint. As loads are all stopped at the same moment and for the same amount of time $T_{dl} - T_n$, they all need to restart together at time $t = 4 + T_{dl} - T_n$. At this moment, all arriving loads are still being idled of precisely $T_{dl} - T_n$. Therefore, at time $t = 4 + T_{dl} - T_n + 1$, the group and the different loads are exactly in the same state than what was observed for the Delay policy. The energy stored by the group experiences the same limits in both Stop and Delay policies. However, the storage process in the Stop policy can be performed with a much higher *power* flexibility (figure 5.12).

$$0 \leq \frac{E_s(t)}{\lambda P_n} \leq T_n(T_{dl} - T_n) \tag{5.27}$$

We denote by $w(t)$ the amount of loads that get idled at time t. Each of such load has a probability $1/T_n$ to have a remaining time ΔT what spreads the recovering of the shifted energy over the next T_n time periods. The relative demand change is expressed as follows.

$$\frac{x(t)}{\lambda P_n} = \left(\frac{1}{T_n} \int_0^{T_n} w(t - \tau) d\tau \right) - w(t) \tag{5.28}$$

In this policy, the action $w(t)$ can be performed on any of the running loads. Its time derivative is therefore not limited.

5.4.3 Local Controller Design

The Stop policy exploits synchronized random number generation (Bernoulli trial). Each running load will decide whether it should be idled on the upcoming time step conducting a Bernoulli trial which success rate is influenced by the frequency deviation at current time. By setting its decision $u^{i,s}(t)$ to 1, load i is idled on $[t, t+1[$.

$$Pr\{u^{i,s}(t) = 1\} = \pi^s(t) = -K_D r^{up}(t) \tag{5.29}$$

Successive load decisions form a stochastic process $u_t^{i,s}$ represented by a random walk with time varying success rate. Principles of stochastic processes should be applied for choosing an appropriate droop K_D that respects the user time deadline with some defined probability. This is what will limit the total FCR capacity that can be provided by the group. The choice of appropriate K_D is very much dependent on the energy content of the reference signal $r^{up}(t)$.

Let's define $T^{run} \in \mathbb{Z}$ as the total load run time. It is the integer-valued time required for a load to consume the required amount of energy E_n when subject to the Stop control policy. It is the sum of the time spent idled and the time spent consuming energy. The time T^{run} is a stopping time w.r.t. the stochastic process $(u_n)_{n\geq 0}$, where each u_k is the result of a Bernoulli trial with success rate $1 + K_D r_k^{up}$ (i.e. probability to run) [154].

$$T^{run} \triangleq \inf \left\{ k \in \mathbb{Z} | u_1 + u_2 + .. + u_k = T_n \right\} \tag{5.30}$$

Let's define γ as the (sufficiently small) probability that, given K_D and the sequence r_k^{up}, the load does not accomplishes its task on time.

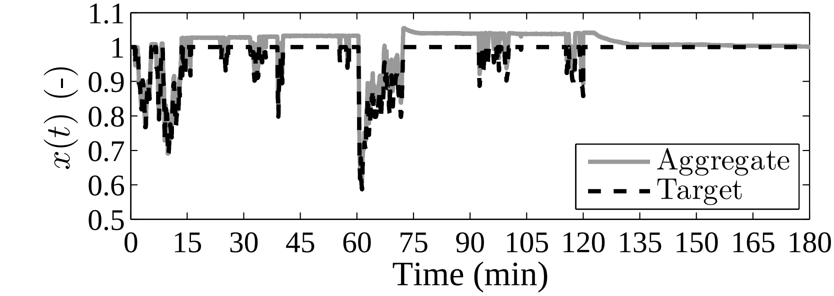
$$Pr\{T^{dyn} > T_{dl} | K_D, r_k^{up}\} = \gamma \tag{5.31}$$

The above expression can be used to define K_D. In addition, hard constraints could be impose such that a load cannot spend too much time being idled (as done in [9]). This will degrade control performances, but to a negligible extend if γ is chosen as sufficiently small.

5.4.4 Frequency response simulation

We conduct the same two simulations than in the previous section. The results are shown on figures 5.13 and 5.14. There are no ramp rate constraint in this case. The demand of the group varies much more rapidly and expresses larger variations (e.g., at t=60min). The energy rebound has a negative impact on the control performance. Yet, the rebound impact is much lower than in the Delay policy, at is occurs with lower magnitude and spans on a larger time period. The longer the initial load run time, the smaller the rebound magnitude.

On both figures, the top charts represent the flexibility $x(t)$ and dotted lines are the corresponding target demand change (neglecting the rebound). The bottom charts highlight the consequences of the Stop control policy on the number of running loads.



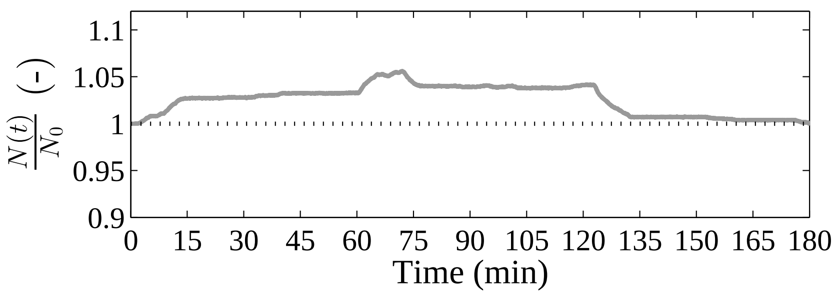


Figure 5.13: Group response to frequency deviations in the Stop control policy. Parameters: $T_n = 1h$, $w_l = T_{dl} - T_n = 30min$, $\lambda = 1s^{-1}$, $P_n = 1kWh$.

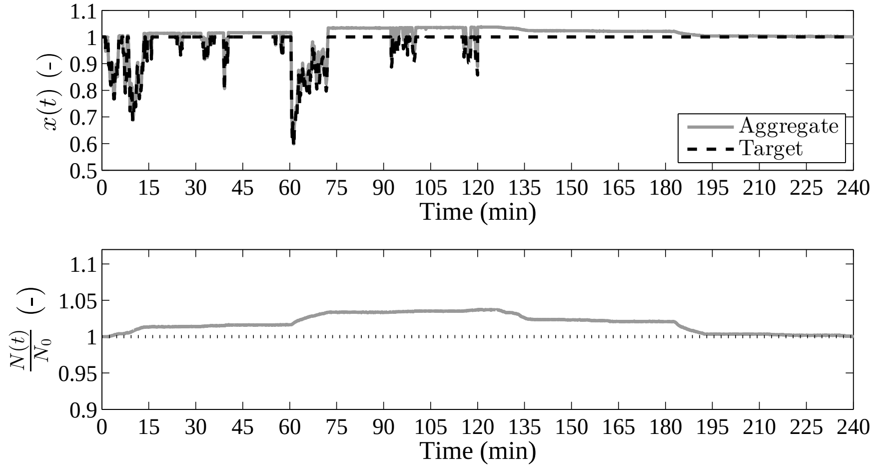


Figure 5.14: Group response to frequency deviations in the Stop control policy. Parameters: $T_n = 1h$, $w_l = T_{dl} - T_n = 30min$, $\lambda = 1s^{-1}$, $P_n = 1kWh$.

5.5 Control Policy 3: power rate modulation

The Rate policy is an even more general form of control. The power of any running load can be continuously set within a certain interval $[P_L, P_H]$. In this chapter, we consider that the load power cannot be increased, $P(t) \in [P_L, P_n]$.

Throughout this work, we've decided to study the rate control policy in its non-selective form. That is, when control decisions are taken, all running loads adapt their power rate in exactly the same proportion. Loads cannot freely choose their power set-point but rather use a fixed proportionality rule w.r.t. the measured frequency deviation $r^{up}(t)$. Therefore the power of each running load, in a homogeneous group, is the same. For convenience, we define the power bias $b(t) \in [-K_D, 0]$ such that,

$$P(t) = P_n\big(1 + b(t)\big) \tag{5.32}$$

Obviously, by neglecting the rebound effect, the relative demand change will follow the reference signal if

$$x(t) \to K_D r^{up}(t) \qquad \Rightarrow \qquad b(t) = K_D r^{up}(t) \tag{5.33}$$

The challenge of this section is to understand the effect of this time varying power rate $P(t)$ on the actual run time of the load. Indeed, the run time of a controlled load is dependent on the power control input. Before entering these discussion, let's firstly conduct at some virtual experiments.

5.5.1 Observation 1: one load changes its power set-point

The unique load arriving at time $t = 4$ is modulated to $P(t) = 2/3kW$. This modulation is imposed (a) during two non-consecutive time steps and (b) during the whole run time of the load. Results are represented on figure 5.15. In this case, the simulation time-step must be reduced in order to capture correctly the effect of the modulation. Indeed, the run time of the modulated load takes non-integer values. On the left-hand charts of figure 5.15, the energy rebound appears at the end of the natural run time ($t = 4 + T_n$). The magnitude of this rebound is equal to the power of the loads at the moment it pops up: $P_n = 1kW$. The rebound lasts for 2/3rd of a second as it must recover the shifted energy ($=2\Delta t \times P_n/3$).

The rebound on the right-hand charts is of lower magnitude. Indeed, the load is still being modulated at the time the rebound appears, at the

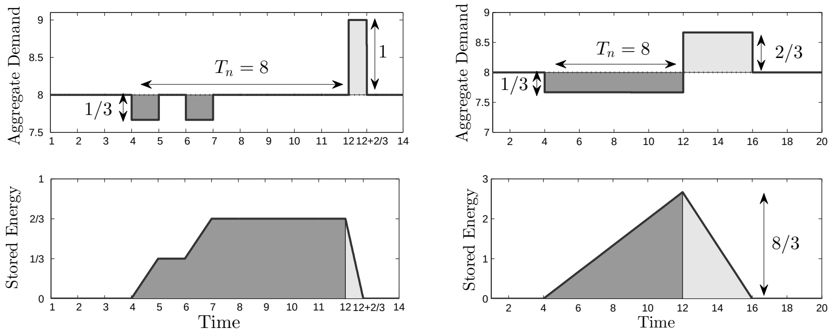


Figure 5.15: Single load impact: power modulation (Rate Policy) during two non-consecutive time-steps (left) and during the whole run time (right).

end of its natural run time.

Consequently, the power demand change $x(t)$ gets influenced by both the current action $b(t)$ as well as by the rebound of past actions $\propto -b(t-\tau)$. In this policy, past action's rebound materializes in a continuous way (more details below). Indeed, as observable on the left-hand chart of figure 5.15, the rebound of the second modulation ($P(t) = 2/3$ at $t = 6$) occurs *right after* the rebound of the first one ($P(t) = 2/3$ at $t = 4$). The time and magnitude of the *first action* rebound depend on the modulation level (i.e. b(t)) at the time it appears. As for $b(t)$, the total time needed to recover the shifted energy is a variable defined on a continuous interval.

5.5.2 Observation 2: modulate all running loads

In a second set of experiments, all running loads are modulated to $P(t) = P_L$ as from $t = 4s$. Results are shown for different group parameters on figure 5.16. In this simulation, we will consider that P_L must be set in order to respect the user deadline in the worst case, that is if $r^{up}(t) = -1, \forall t$. This is overly restrictive, as shown below.

On 5.16, the group demand is step-wise constant as loads arrivals are concentrated on discrete time steps in the performed simulation (and so is it for their stopping time).

Due to our choice of P_L, the stored energy is lower than in the two previous policies. Indeed, as load cannot be completely stopped, the rate at which they postpone their energy consumption is limited. Similarly to the two other control policies, loads begin to recover the non-consumed energy

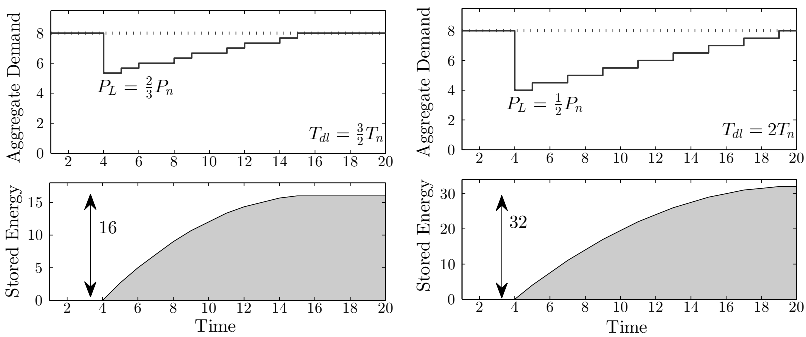


Figure 5.16: Power rate change for all running and arriving loads to $P(t) = P_L$ from $t = 4s$ until they individually reach their energy level. Parameters: $T_n = 8s$, $P_n = 1kW$, $\lambda = 1s^{-1}$. Left chart: $T_{dl} = 12s$. Right chart: $T_{dl} = 16s$.

exactly T_n instants after their arrival time t_a. However the amount of energy they were actually able to postpone in the interval $[t_a, t_a + T_n]$ is bounded by $(P_n - P_L)$. Altogether, loads with the same user deadline T_{dl} controlled in the Rate policy can shift half of the energy that could have been shifted with the Stop policy.

$$0 \leq \frac{E_s(t)}{\lambda P_n} \leq \frac{1}{2}T_n(T_{dl} - T_n) \tag{5.34}$$

5.5.3 Local control design

As was described above, the rate decision is deterministic. It corresponds to define the bias $b(t)$.

$$u^{i,r}(t) = b(t) = K_D r^{up}(t) \tag{5.35}$$

Setting K_D such that $1 - K_D P_n = P_L$ with the above defined P_L leads to the following.

$$P_L T_{dl} = (1 - K_D)P_n T_{dl} \geq E_n \quad \Rightarrow \quad K_D \leq \frac{T_{dl} - T_n}{T_n} \tag{5.36}$$

Yet, this can be overly restrictive. Loads may technically vary their power rate in a larger range. The ideal K_D much compromises between the technical capability and the user deadline. As was discussed for the Stop Policy, K_D must be selected such that the maximum run time of loads respects the user time deadline. This is very much dependent on the energy content of the reference signal, as discussed below.

Energy-time constraint and its consequences

We would like to model the continuous nature of the energy rebound. As mentioned above, loads are subject to an energy constraint and a time deadline constraint. Furthermore, all loads are supposed to have exactly the same power rate $P(t)$. Therefore, the energy and time constraints can be combined in a very general way (eq. (5.37)).

$$\int_0^{T_{dl}} P(t-\tau)\,d\tau \geq E_n \qquad \forall t \tag{5.37}$$

Stopping Time of a single load

The stopping time $t_s(P(t), t_a)$ is the instant in time at which a load has consumed E_n given its arrival time t_a and the power input $P(t)$.

$$t_s(P(t), t_a) \triangleq \min\{t > t_a | \int_{t_a}^{t} P(\tau)\,d\tau = E_n\} \forall t \in [t_a, t_s] \tag{5.38}$$

This stopping time is directly dependent on the actual realization of power $P(t)$. It introduces a strong non-linearity in the model : the upper bound of the integral is implicitly defined.

Run Time of a single load

Knowing the instant in time $t_s(P(t), t_a)$ at which the load stops, it is straightforward to compute its exact run time T^{run}. The run time is the period of time during which the load was consuming energy.

$$T^{run}(P(t), t_a) = t_s(P(t), t_a) - t_a \tag{5.39}$$

As (5.35) gives $b(t) \in [-K_D, 0]$, the run time is bounded to $T^{run} \in \mathcal{T} = [T_n, T_n/(1-K_D)]$. A formal definition is found below.

$$T^{run}(P(t), t_a) \triangleq \min\{\theta \in \mathcal{T} | \int_0^{\theta} P(t_a+\tau)\,d\tau = E_n\} \quad \forall t \in [t_a, t_s] \tag{5.40}$$

Dynamic Run Time as a group variable

The above run time T^{run} was defined for a single load, starting at $t = t_a$ and subject to power control $P(t)\,\forall t \in [t_a, t_s]$. It is easily generalized to any load of the group. Assuming that the arrival process can be represented in by a continuous starting density, some loads are starting (with density λ) and some other are stopping at every moment in time t. The dynamic run time $T^{dyn}(P(t))$ is the run time of loads *stopping* at time t and subject to control inputs $P(t)$.

$$T^{dyn}(P(t)) \triangleq \min\{\theta \in \mathcal{T} | \int_0^{\theta} P(t-\tau)\, d\tau = E_n\} \quad \forall t \tag{5.41}$$

Equation (5.41) is defined *backward* w.r.t time t. It is a causal definition: we search the run time of the loads that are stopping at time t looking at the past modulation they have just been subject to.

Time distortion and rebound

The above expression is equivalent to a *time schedule distortion.*

$$T_n = \int_0^{T^{dyn}(t)} \frac{P(t-\tau)}{P_n}\, d\tau = T^{dyn}(t) + \int_0^{T^{dyn}(t)} b(t-\tau)\, d\tau \tag{5.42}$$

The power modulation acts as if a virtual time distorting effect was continuously *reshaping* the initial consumption profile. An increase in the consumption time ($T^{dyn}(t) > T_n$) leads to a positive *energy pay-back* or rebound (i.e., demand is higher than expected), and conversely.

Impact on the power demand

Using the above definition, the power of a group of loads subject to power modulation is found below.

$$N(t) = \int_0^{T^{dyn}(t)} S(t-\tau)d\tau = \lambda T^{dyn}(t) \tag{5.43}$$

$$D(t) = \lambda P_n(1+b(t))T^{dyn}(t) \tag{5.44}$$

The relative demand change $x(t)$ is as follows.

$$\begin{aligned}
x(t) &= \frac{1}{\lambda P_n T_n}\Big(\lambda P_n(1+b(t))T^{dyn}(t) - \lambda P_n T_n\Big) && (5.45)\\
&= b(t)\frac{T^{dyn}(t)}{T_n} - \frac{1}{T_n}\int_0^{T^{dyn}(t)} b(t-\tau)\, d\tau && (5.46)\\
&= b(t) - \frac{1+b(t)}{T_n}\int_0^{T^{dyn}(t)} b(t-\tau)\, d\tau && (5.47)\\
&= x^{FCR}(t) - e_D(t) && (5.48)
\end{aligned}$$

The target part of the demand change is $x^{FCR}(t) = b(t)$ to which an error term $e_D(t)$ is added, as a consequence of the energy rebound.

$$e_D(t) = \frac{1+b(t)}{T_n}\int_0^{T^{dyn}(t)} b(t-\tau)\, d\tau \tag{5.49}$$

Let's analyze this error term. The error corresponds to the power demand of all additional running loads $(N(t)-N_0)$. Their (relative) power rate is equal to $P(t)/P_n = (1+b(t))$, and their exact number is a consequence of the *average* energy that was shifted in the $T^{dyn}(t)$ previous instants, shown below.

$$\frac{1}{T^{dyn}(t)} \int_0^{T^{dyn}(t)} b(t-\tau)\,d\tau \tag{5.50}$$

As can be seen, this is not what appears in the error term. Indeed, the recovered energy get's *concentrated* at the end of the natural run time of each load. Indeed, as observed on figure 5.15, the time at which the rebound of a certain action occurs is dependent on the remaining time of the load at the moment the action is taken. This remaining time can be anywhere between 0 and T_n. This can intuitively explain why the energy content is actually divided by the probability of seeing a certain remaining time $1/T_n$.

$$\frac{1}{T_n} \int_0^{T^{dyn}(t)} b(t-\tau)\,d\tau \tag{5.51}$$

The behavior of the group subject to such control policy are explored in more details in the next chapter. In the rest of this chapter, we will show that, if the energy content of $b(t)$ is small enough, the demand change $x(t)$ is approximately equal to the one in developed in the Stop Strategy.

5.5.4 Frequency response simulation

The same frequency tracking simulations are conducted. Their results are shown on figures 5.17 and 5.18. These figures are almost exactly similar to the one presented in the Stop Strategy (figures 5.13 and 5.14). More detailed comparison are done in the following sections.

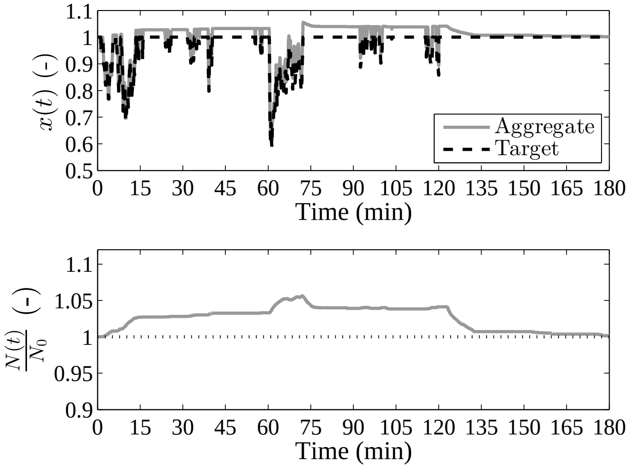


Figure 5.17: Group response to frequency deviations in the Rate control policy. Parameters: $T_n = 1h$, $w_l = T_{dl} - T_n = 30min$, $\lambda = 1s^{-1}$, $P_n = 1kWh$.

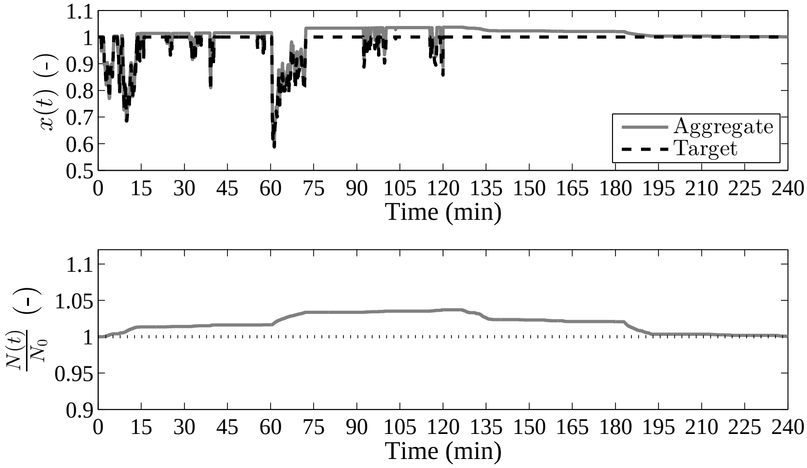


Figure 5.18: Group response to frequency deviations in the Rate control policy. Parameters: $T_n = 1h$, $w_l = T_{dl} - T_n = 30min$, $\lambda = 1s^{-1}$, $P_n = 1kWh$.

5.6 Aggregate models: from homogeneous to heterogeneous groups

The models presented above are relative to an homogeneous group with parameters $S(t) = \lambda$, $\mathcal{L} : \{P_n, T_n, E_n, T_{dl}\}$. They are presented below in a condensed form. Superscripts $j = \{d, r, s\}$ refer to the three policies: (d) Delay, (s) Stop, (r) Rate.

Aggregate models in homogeneous groups

These aggregate models represent the controlled demand of an homogeneous ECL group providing FCR with droop K_D in each considered policy.

$$\hat{x}^d(t|T_n) = K_D\Big(\tilde{r}^{up}(t, T_n) - \tilde{r}^{up}(t - T_n, T_n)\Big) \tag{5.52}$$

$$\hat{x}^s(t|T_n) = K_D\left(r^{up}(t) - \frac{1}{T_n}\int_0^{T_n} r^{up}(t-s)ds\right) \tag{5.53}$$

$$\hat{x}^r(t|T_n) = K_D\left(r^{up}(t) - \frac{1 + r^{up}(t)}{T_n}\int_0^{T^{dyn}(t)} r^{up}(t-s)ds\right) \tag{5.54}$$

The objective of this section is to find an equivalent formulation for heterogeneous groups of ECL, which parameters are distributed on $P \in [P_m, P_M]$ and $T \in [T_m, T_M]$ with probability $\Omega_{P,T}$. The heterogeneous group can be seen as a weighted sum of homogeneous groups of parameter $T \in [T_m, T_M]$ (e.g.,[3]). The contribution of each homogeneous group should be weighted by the ratio of the number of running loads $N_{0|T}/N_0 = T/\mathbb{E}[T]$ and the marginal distribution of parameter T, denoted Ω_T. Indeed, homogeneous models $\hat{x}^j(t|T)$ relative to each strategy j are computed for groups counting $N_{0|T} = \lambda T$ running loads, while the heterogeneous groups hold on average $N_0 = \lambda\mathbb{E}[T]$ running loads. The resulting heterogeneous model is independent of the random power rate P.

$$\hat{x}^j(t) = \int_{T_m}^{T_M} \hat{x}^j(t|T)\frac{N_{0|T}}{N_0}\Omega_T\, dT = \frac{1}{\mathbb{E}[T]}\int_{T_m}^{T_M} \hat{x}^j(t|T)T\Omega_T\, dT \tag{5.55}$$

where $\mathbb{E}[T] = \int_{T_m}^{T_M}\int_{P_m}^{P_M} T\,\Omega_{P,T}\, dP\, dT$, and $\Omega_T = \int_{P_m}^{P_M}\Omega_{P,T}\, dP$.

The Delay and Rate aggregate models are more computationally demanding when applied to simulations. The algorithm will need to keep track of additional *run-time-specific* elements than in the Stop case: (1)

Aggregate models in heterogeneous groups

The following aggregate models are relative to heterogeneous ECL groups providing FCR with droop K_D in each considered policy, given distribution Ω_T.

$$\hat{x}^d(t) = \frac{K_D}{\mathbb{E}[T]} \int_{T_m}^{T_M} \big(\tilde{r}^{up}(t,T) - \tilde{r}^{up}(t-T,T)\big) T \Omega_T \, dT \tag{5.56}$$

$$\hat{x}^s(t) = K_D \bigg(r^{up}(t) - \frac{1}{\mathbb{E}[T]} \int_{T_m}^{T_M} \Omega_T \int_0^T r^{up}(t-s) ds \, dT \bigg) \tag{5.57}$$

$$\hat{x}^r(t) = K_D \bigg(r^{up}(t) - \frac{1 + r^{up}(t)}{\mathbb{E}[T]} \int_{T_m}^{T_M} \Omega_T \int_0^{T^{dyn}(t,T)} r^{up}(t-s) ds \, dT \bigg) \tag{5.58}$$

the limited ramp rate represented by the filtered signal $\tilde{r}^{up}(t,T)$ in the Delay case and (2) the non-explicit computation of the dynamic run time $T^{dyn}(t,T)$ in the Rate case.

Some illustrative examples are shown below, where the group's parameters P and T are distributed differently.

1. Independent and Uniformly distributed (Fig. 5.19 and 5.20).

$$\Omega_{P,T} = \frac{1}{(P_M - P_m)(T_M - T_m)} \qquad \forall P \in [P_m, P_M], \forall T \in [T_m, T_M]$$

2. Linearly dependent on T and uniformly distributed. We define $P_M(T) = P_m + \frac{T - T_m}{T_M - T_m}(P_M - P_m)$ as the maximum observable power of loads with run time T (Fig. 5.21 and 5.22).

$$\Omega_{P,T} = \begin{cases} \dfrac{1}{(P_M(T) - P_m)(T_M - T_m)} & , \forall P \in [P_m, P_M(T)], \forall T \in [T_m, T_M] \\ 0 & \text{elsewhere} \end{cases}$$

The second case may seem a little tricky, but is designed to illustrate that heterogeneous models are independent of P, if the marginal distribution Ω_T is correctly computed. In this example, the distribution is made such that loads running for short time tend to have lower power rate.

In all figures presented below (5.19,5.20, 5.21 and 5.21), the dashed lines represent the heterogeneous aggregate model and the plain lines are individual simulations where each load is separately modeled. Both match very well with each other, as long as the parameter distributions are known. The reasons explaining the difference between the two curves is simple: the plain lines are random. They necessarily variate around their expected value represented by the dotted lines (aggregate model). More details on the model accuracy are given in the following section.

Figures 5.19 and 5.20 show the FCR response of the group, controlled in the three considered policies, for loads with relatively small run time and considering the two above described distributions (independent uniform and linear dependent). Figures 5.19 and 5.20 are relative to loads with longer run times. Here is what can already been observed, in terms of group dynamics.

In general, the longer the load's run time, the smoother the rebound. This is valid for all control policies.

In the Delay policy, the parameter randomness allows to smooth out the rebound impact. As can be seen from comparison of (C.1) and (5.56), the rebound impact in the heterogeneous case ($r^{up}(t-T,T)$) is averaged out by the run time distribution. This is clearly visible by comparing *relative* rebound magnitude w.r.t. the initial demand change on e.g., of the top charts of figures 5.19 and 5.21. Let's also note the impact of the ramp rate limit: the efforts deployed by group with longer run times (e.g., 5.21) are largely below the ones that are observed for shorter run times (e.g., 5.19).

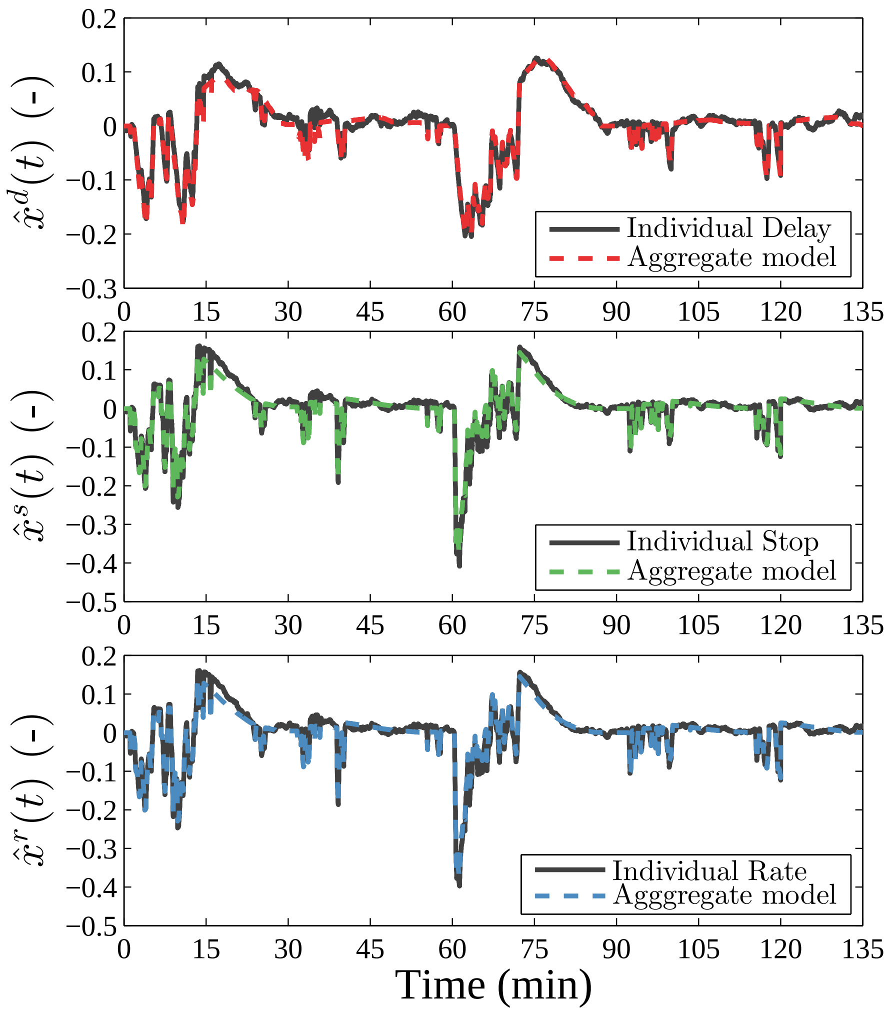


Figure 5.19: Heterogeneous group of loads with short run time with autonomous power frequency control (parameters are independent uniformly distributed). Plain lines: individual model (a group where each load is individually simulated). Dashed Lines: aggregate model. Parameters: $T \in [100, 900]s$, $P \in [1, 10]kW$, $Tdl/T_n = K_D = 50\%$, $\lambda = 10s^{-1}$.

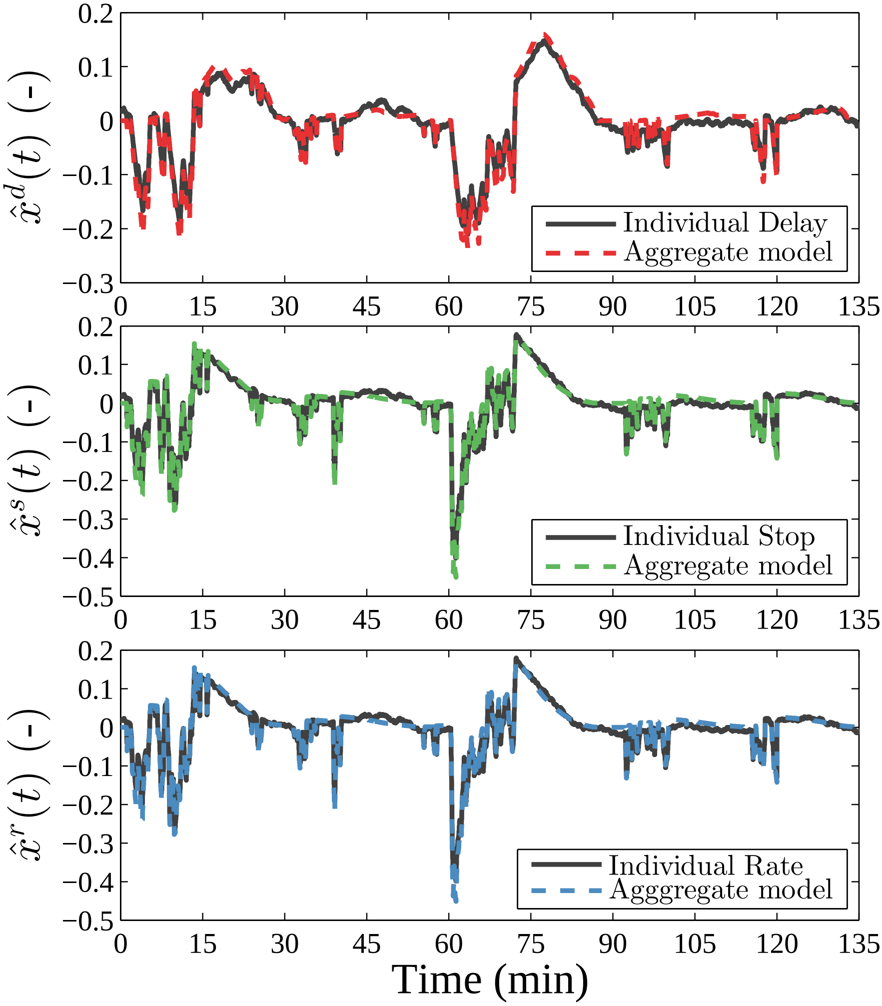


Figure 5.20: Heterogeneous group of loads with short run time with autonomous power frequency control (parameters are linearly dependent uniformly distributed). Plain lines: individual model (a group where each load is individually simulated). Dashed Lines: aggregate model. Parameters: $T \in [100, 900]s$, $P \in [1, 10]kW$, $Tdl/T_n = K_D = 50\%$, $\lambda = 10s^{-1}$.

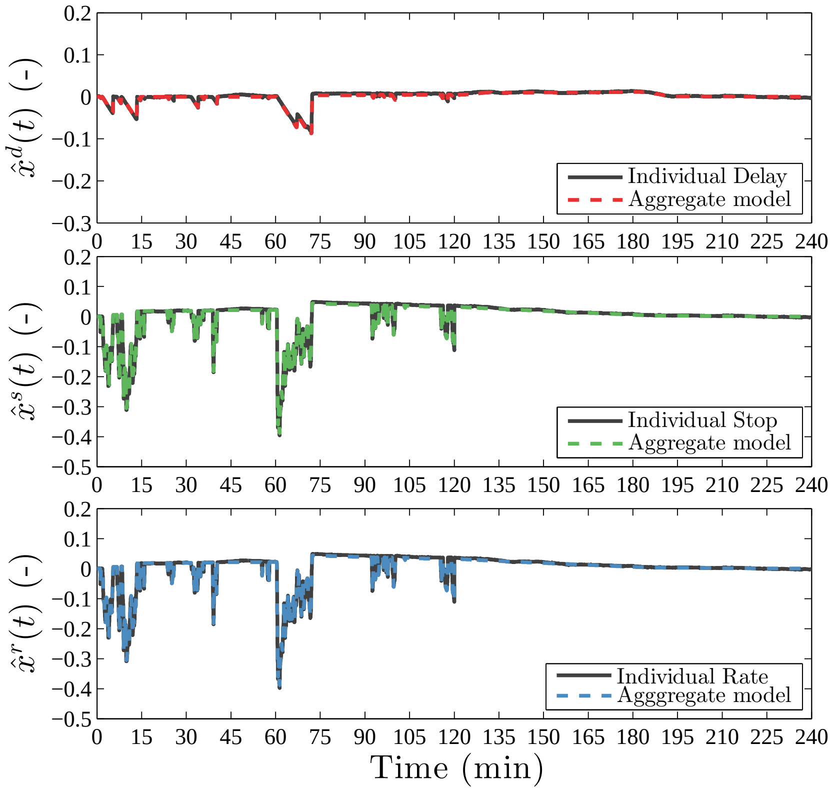


Figure 5.21: Heterogeneous group of loads with long run time with autonomous power frequency control (parameters are independent uniformly distributed). Plain lines: individual model (a group where each load is individually simulated). Dashed Lines: aggregate model. Parameters: $T \in [1,2]h$, $P \in [1,10]kW$, $Tdl/T_n = K_D = 50\%$, $\lambda = 10s^{-1}$.

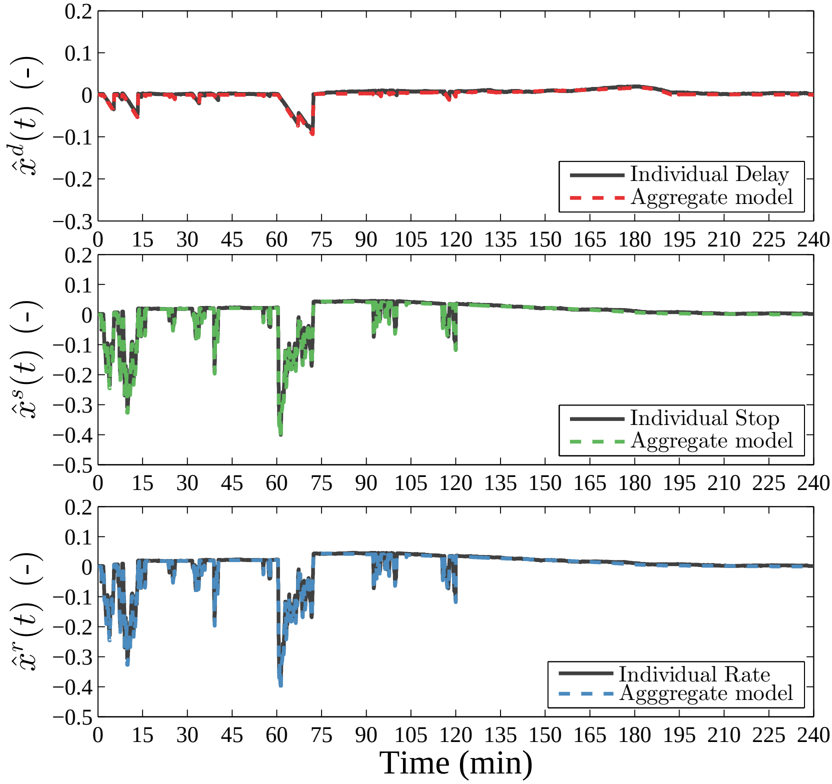


Figure 5.22: Heterogeneous group of loads with long run time with autonomous power frequency control (parameters are linearly dependent uniformly distributed). Plain lines: individual model (a group where each load is individually simulated). Dashed Lines: aggregate model. Parameters: $T \in [1,2]h$, $P \in [1,10]kW$, $Tdl/T_n = K_D = 50\%$, $\lambda = 10s^{-1}$.

5.7 Model Accuracy in different load scenarios

The aggregate models accuracy must be assessed by comparing their outcome to detailed simulations where loads are separately modeled. Two scenarios are explored (Table 5.2). Scenario I exploits common types of residential ECLs with a power typically below 2kW and running for a quarter to half an hour (e.g., Water heating in white appliances [90]). Scenario II aggregates larger loads such as electric vehicles or night storage heaters which power is between 2 and 7 kW and running duration from 15 minutes to 3 hours [112].

Table 5.2: Load parameters in the two scenarios ($K_D = T_{dl}/T_n = 30\%$).

	Name	Value	Name	Value
Scenario I	Duration	$T \in [1/4, 1/2]$ h	Power	$P \in [1, 2]$ kW
Scenario II	Duration	$T \in [1/4, 3]$ h	Power	$P \in [2, 7]$ kW

In appendix C, the interested reader can find detailed evaluation of an approximation of the aggregate model in the rate policy. For large run time T_n and relatively small energy content of the reference signal $r^{up}(t)$ in any interval $[t - T_n, t]$ (which is the case in practice), we show that the following approximation is valid.

$$\hat{x}^s(t) \simeq \hat{x}^r(t) \qquad \text{for large } T_n \tag{5.59}$$

We will use this approximation in the rest this work. In order to empirically validate the use of all aggregate models, we conduct a Monte Carlo analysis to evaluate the following error terms in the three control policies.

$$e_i^j(t) = \hat{x}^j(t) - x_i^j(t) \qquad \forall j = \{d, s, r\} \tag{5.60}$$

We compare the exact individual simulations $x_i^j(t) \forall i = 1..N_{sim}$ in each policy j to its aggregate estimate $\hat{x}^j(t)$. In the rate policy, let's note that we use the above mentioned approximation and deliberately compare the individual model $x^r(t)$ simulated with the Rate control policy to the aggregate estimate relative to the Stop policy $\hat{x}^s(t)$. Both estimation and approximation errors are estimated together in this case. If the error is small enough, it will show that the Rate aggregate model is suited to represent both Rate and Stop policies.

5.7.1 Accuracy metric

In practice, an FCR provider must guarantee its response to frequency deviations stays within a tolerance band of ±5% around the frequency-based set-point. This can be deduced from the ±10mHz tolerance, in article 44 of the Network Code on Load-Frequency Control [52], and that is referred to as the *Maximum combined effect of inherent Frequency Response Insensitivity and possible intentional Frequency Response Dead band of the governor of the FCR Providing Units or FCR Providing Groups*. Let's note that the tolerance is larger in, e.g., the UK (±7.5%).

In the following chapter of our work, the aggregate models will be used in long-term simulations to assess such response performance and to conduct economic analyses. Therefore, those models must have a very good accuracy, leading their estimation error to be negligible compared to the 5% ENTSO-E tolerance. The normalized root-mean-square estimation error derived from each e_t^j should stay below 0.5% and the maximum normalized estimation error below 1%.

5.7.2 Independent uniform distributions

In a first set of simulations, load parameters joint distribution $\Omega_{P,T}$ is assumed to be the product of two independent uniform distributions. A 4-hour simulation is run in the two scenarios for $N_{sim} = 50$ realizations of the group's demand and for 4 different group sizes (λT_n). Normalized root-mean-square error distribution and maximum error distributions are shown respectively on Fig.5.23 and Fig.5.24 (with Table 5.3). For comparison purpose, the dashed lines on Fig.5.23 represent the volatility v_D (eq. (4.59), Poisson arrivals).

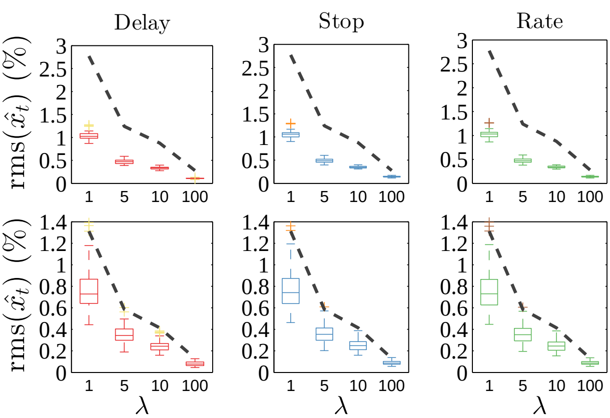


Figure 5.23: Distributions of the root mean square error of the estimate with varying group size. Top: Scenario I. Bottom: Scenario II. Dashed: volatility interval v_D. Parameters are independent and uniformly distributed.

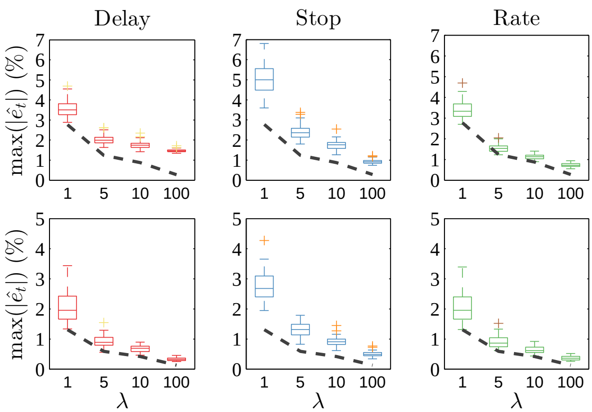


Figure 5.24: Distributions of the maximum error of the estimate with varying group size. Top: Scenario I. Bottom: Scenario II. Dashed: volatility v_D. Parameters are independent and uniformly distributed.

In these simulations, the arrivals are constant. Interestingly though, the maximum observed absolute error corresponds more or less to $3v_D$, which was the choice made for the volatility interval. The results shown on table 5.3 indicates that the models are perfectly suited to represent groups whose arrival rate is situated between 10 and 100 $[s^{-1}]$.

Let's note that the maximum observed absolute estimate error tends to be higher in the Stop policy than in the Rate policy. This comes from the *selective* nature of the Stop policy that adds an extra layer of uncertainty: loads are randomly selected to get switched on or off.

Table 5.3: Maximum observed error (in %).

Policy	Delay			Stop			Rate		
$\lambda[s^{-1}]$	5	10	100	5	10	100	5	10	100
Scenario I	3.2	2.5	2	3.3	3	1	2	1.5	1
Scenario II	2.3	2	1.7	2	1.6	0.7	1.3	0.9	0.6

Monte Carlo Convergence analysis could have been performed for additional precision, but our goal is to highlight the order of magnitudes.

5.7.3 Linearly dependent uniform distributions.

For the sake of generality, a second set of simulations is considered in which a linear dependence exists between the maximum power rate $P_M(T)$ and the actual run time of the load T. Normalized root-mean-square error distribution and maximum error distributions are shown respectively on Fig.5.25 and Fig. 5.26. In comparison to the previous case, the error is in the same order of magnitude. The prior knowledge of the exact parameter distribution is absolutely crucial to get to such low error levels.

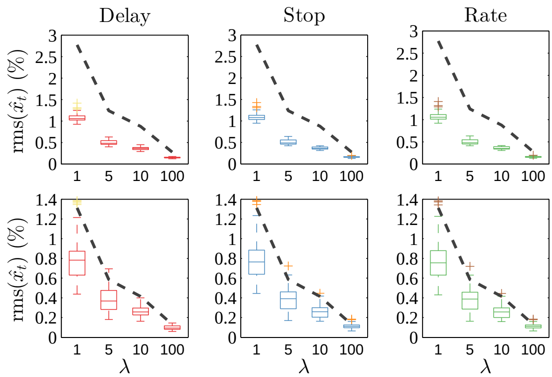


Figure 5.25: Distributions of the root mean square error of the estimate with varying group size. Top: Scenario I. Bottom: Scenario II. Dashed: volatility interval v_D. Parameters are uniformly distributed with linear dependence.

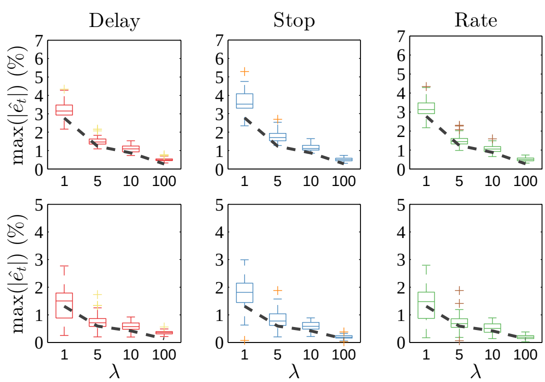


Figure 5.26: Distributions of the maximum error of the estimate with varying group size. Top: Scenario I. Bottom: Scenario II. Dashed: volatility interval v_D. Parameters are uniformly distributed with linear dependence.

5.8 Chapter conclusion

The local controller design approach that was undertaken in this chapter can be summed up in the following process.

1. Define the incremental control action (i.e., delay, stop, rate).
2. Observe its impact on the relative group demand change, also known as *relative flexibility* $x(t) = D(t)/D_0 - 1$.
3. *Ignore the rebound*, and deduce the *number* of such actions necessary to impose a predefined change to the group demand (i.e., FCR control $x(t) \rightarrow K_D r^{up}(t)$).
4. Find a way to compute this number of action in an autonomous fashion, based on local information (e.g., $\Delta f(t), T$).

The third element is crucial: in this chapter loads do not counteract the rebound impact in any way. This will be the focus of the last chapter of the present work.

After this design process, the expected reaction of homogeneous and heterogeneous group has been condensed into aggregate models. These are empirically shown to accurately model the group's response, as long as the group size is sufficient. Most of all, they will prove very useful to run long-term simulations and estimate the year-long and system-wide impact of exploiting ECL within FCR.

Chapter 6

The economics of autonomous frequency control

Chapter summary

Bud Fox: How much is enough?

Gordon Gekko: It's not a question of enough, pal. It's a zero sum game, somebody wins, somebody loses. Money itself isn't lost or made, it's simply transferred from one perception to another. [...] It's all about bucks, kid. The rest is conversation.

G. Gekko (Michael Douglas), B.Fox (Charlie Sheen), *Wall Street* (1987)

In this chapter, the costs and benefits of autonomous power control are assessed. Our objective is to simulate the overall system impact of a massive deployment program in which a very large number of small electric appliances are equipped with one of the above discussed controllers. Large groups of flexible loads replace part of the generation assets providing Frequency Containment Reserve.

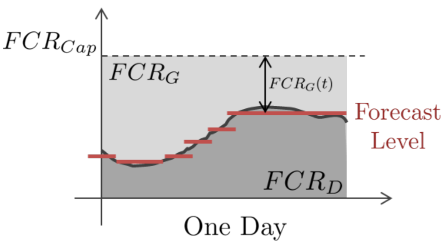


Figure 6.1: Adapting to the available demand-side response.

Practically, large groups of small electric appliances will have a time varying power consumption (e.g., EVs [73]). The vision developed throughout this study is that the system operator will adapt to such variation. The process is illustrated on figure 6.1 where different level of FCR are sourced from the generation-side at different moment in time. The system requires a certain total capacity FCR_{cap} in order to guarantee the system stability as well as frequency quality. The operator will source a first part of these capacity needs from the available demand-side reserve $FCR_D(t)$ and source the remaining part from the generation-side $FCR_G(t)$. The same concept applies to seasonal or daily variations.

In the rest of this chapter, we will however consider an ideal case in which the provided capacity is constant. This leads to an overestimate of potential benefits, but has the advantage of being more simple and transparent. In order to evaluate the overall system impact of exploiting ECL as FCR providers, a simple one-bus power system model is presented below. It will be used to represent the system frequency dynamics. The model is adapted to integrate different shares of FCR provided by loads. Firstly, it is used in a short-term, event-based simulation. This allows to evaluate the frequency dynamics following a major incident (N-2 criterion: loss of two large generation assets). The model is later exploited in long term historical simulations. An idealized view of the Continental Europe power system reconstructs historical frequency deviations and FCR activation levels. Several simplifying assumptions are used but the important *trends* are very well highlighted by this approach.

6.1 One-bus Power System Model

The power system frequency excursion $\Delta f(t) = f(t) - f_n$ at time t directly depends on the *system imbalance* $I(t)$, the difference between total system generation and demand. Frequency dynamics are represented by the following equation (adapted from [6, 111]).

$$M\frac{df(t)}{dt} = I(t) - (K_{FCR} + L)\Delta f(t) - \frac{K_{aFRR}}{\tau_{aFRR}}\int \Delta f(t)\,dt - D_0 x(t) \qquad (6.1)$$

$I(t)$	MW	System power imbalance at time t.
M	MWs/Hz	Angular Momentum of the rotating masses.
$-K_{FCR}$	MW/Hz	Negative proportional feedback (FCR).
$-L$	MW/Hz	Natural frequency-dependence of total system load.
K_{aFRR}	MW/Hz	Gain of integral feedback (aFRR).
τ_{aFRR}	s	Time constant of integral feedback (aFRR).
$D(t)$	MW	Controlled power demand of the ECLs group.
D_0	MW	Initial (Baseline) demand of the group.
$x(t)$	MW	Relative flexibility $D(t)/D_0 - 1$ (aggregate model).

$$\Delta P_g^{set}(t) \longrightarrow \boxed{\frac{1}{\tau_g s + 1}} \longrightarrow \Delta P_g(t)$$

Figure 6.2: Transfer function of generation (Laplace domain).

The model considers the response of FCR (proportional with gain K_{FCR}), aFRR or secondary control (integral with gain K_{aFRR} and time constant τ_{aFRR}), as well as the contribution of the controlled loads $D_0 x(t)$. In addition, natural stabilization effect provided by total system load (L) is taken into account. In addition to Eq.(6.1), the finite response-time of generators is modeled by a first-order transfer function (Fig. 6.2). The time constant τ_g is set to $5s$. We introduce the integral action to highlight the impact of the rebound on slower reserve use.

6.2 Performance Requirements

The performance that must be guaranteed by an FCR provider vary from one system to the other. In Europe, ENTSO-E is in process of harmonizing these requirements. At this stage (end 2015), the different system operators can still specify their own requirements and the verification procedure used for ex-post performance assessment. Several elements are set in common in the *network code for load-frequency control and reserve* [52]. The FCR participants must always guarantee that the deployed flexible capacity stays in a ±10mHz band around the set-point defined by the power-frequency curve. This is illustrated on figure 6.3.

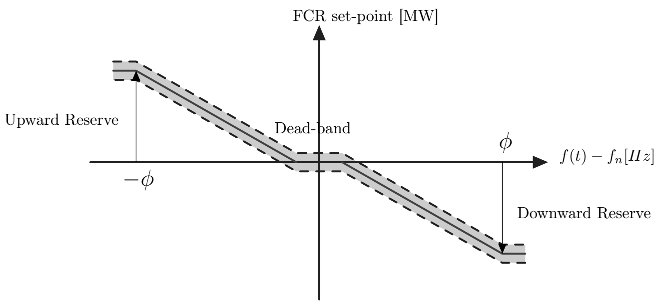


Figure 6.3: Power-frequency curve defining FCR's set-point and tolerance band.

In ENTSO-E systems, the frequency control is designed such that steady-state frequency deviations are contained in an interval $\Delta f(t) \in [-\phi, \phi]$. Furthermore, the FCR response saturates for frequency values that goes out of this band. In Continental Europe, $\phi = 200$mHz. Therefore, at the extreme points, the ±10mHz tolerance corresponds to a ±10mHz/200mHz= ±5% tolerance. This means that the power that is actually deployed by the FCR participant must stay in a ±5% tolerance band around its frequency-based set-point. This requirement is a guidance principle. In fact, larger tolerance is accepted by the system operators. The Inter-TSO platform *regelleistung* mentions a tolerance of +20% *above* the set-point value is accepted [148] in FCR prequalification[1] documents.

In opposition to ENTSO-E requirements, a resource that can only guarantee a ±10% control accuracy could potentially participate in the German system. This resource could indeed offer 90% of its available flexibility in such market and then act as if 100% had been offered. Let's note however

[1] `https://www.regelleistung.net/ext/static/prequalification`: used by the four TSO's in Germany and partly by other TSO in the Netherlands, Austria, etc.

that the prequalification test themselves [115] are much stricter in Germany than, e.g., in Belgium. Indeed, they require to track with very large precision an activation profiles that has a relatively high energy content.

For a group of ECL delivering upward FCR to have acceptable performances, the flexibility $x(t)D_0$ offered by the group, including all rebound errors, will need to stay within a small interval around the frequency-dependent set-point. This is illustrated on figure 6.4.

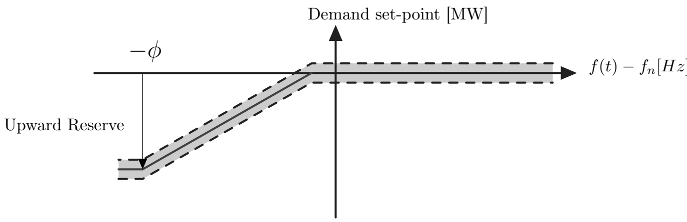


Figure 6.4: FCR set-point and tolerance band for delivering upward Reserve.

6.3 Event-based Simulation of FCR.

In this section, the simulated system is subject to a sudden and major imbalance. The system is initially in steady-state ($\Delta f(t=0) = 0, D(t=0) = D_0$). The situation is representative of losing two large generators in Continental Europe (N-2 criterion) together representing 3000MW. System parameters are extracted from NC-LFCR [52].

Base Case: no demand response (DR) participation.

The Base Case (no DR) simulation parameters are shown on Table 6.1.

Table 6.1: Parameters of the system in the Base Case simulation.

Name	Value	Name	Value	Name	Value
I	-3000 MW	M	25 GWs/Hz	τ_{aFRR}	100 s
$K_{FCR} = K_{aFRR}$	15 GW/Hz	L	4 GW/Hz	τ_g	5 s

Case Studies: FCR provision from ECLs of different kind.

A large homogeneous group of ECLs now provides the entire FCR capacity required at system level. The group is represented by the Rate/Stop aggregate model $x^r(t|T_n)$.

$$x^r(t|T_n) = \frac{K_D}{\phi}\left(\Delta f(t) - \frac{1}{T_n}\int_0^{T_n} \Delta f(t-s)\, ds\right) \tag{6.2}$$

Four tests cases are considered in which the run time parameter T_n of the participating loads is successively set to $1/4h$, $1/2h$, $1h$ and $2h$. The provided FCR capacity $K_D D_0 = 3000MW$ is constant. We fix K_D to 50% and therefore D_0 to $6000MW$. The parameter product λP_n is adapted accordingly in all cases. The response time of demand-side resources is neglected. The group is assumed able to deploy its full capacity within one time-step (one second). Simulation parameters are shown on Table 6.2.

Table 6.2: Parameters for the simulated cases with active loads.

Name	Value	Name	Value	Name	Value
I	-3000 MW	M	25 GWs/Hz	K_D	**0.5**
D_0	**6000 MW**	L	4 GW/Hz	ϕ	0.2 Hz
K_{FCR}	**0 GW/Hz**	K_{aFRR}	15 GW/Hz	τ_{aFRR}	100 s

6.3.1 Simulation Results

Immediate contribution of loads: improved stability.
As shown on figure 6.5 (top chart), FCR volumes sourced from traditional generation ($= FCR_G$) have been completely replaced by the active loads in the case studies.

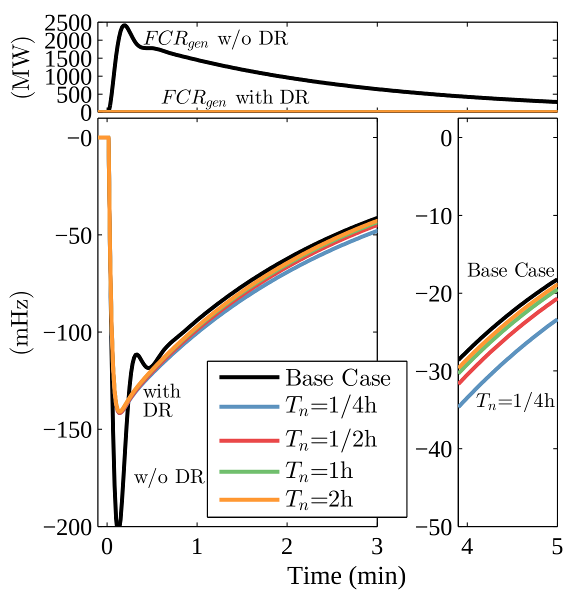


Figure 6.5: FCR from generation (Top) and frequency change Δf_t (Bottom) evolutions during the five first minutes following the event (occuring at t=0).

The frequency excursions are contained in a tighter interval ($\Delta f >$-150mHz) when FCR are provided by loads. They reach down to -200mHz in the base case scenario. This increased frequency quality comes from the ability of loads to respond in a much faster way than generators to frequency changes. The short-term stability of the system seems increased.

Furthermore, as highlighted on the left chart of Fig. 6.6, the contribution of loads to primary control diminishes the total volumes of FCR that are required. In the base case (dotted black curve), generation reacts more slowly to frequency changes. About 80% of the available capacity of FCR_G are activated. In the other four studied cases, the active loads groups respond quickly by consuming less power (they provide positive reserve). Only 70% of the available FCR capacity (i.e. $K_D D_0$) is deployed.

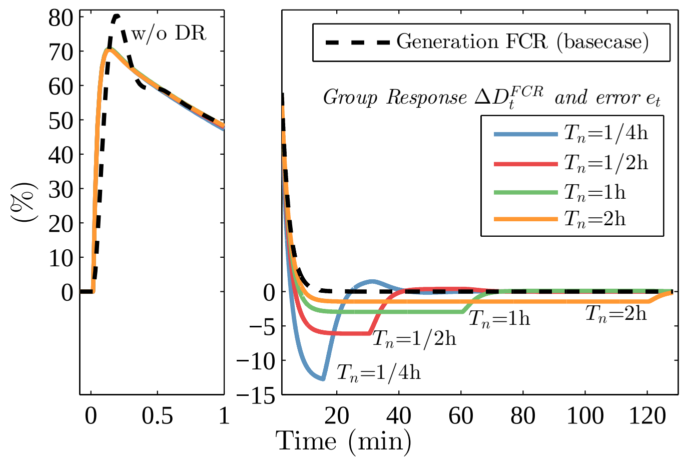


Figure 6.6: FCR activation in % of the offered capacity (3GW). Dotted line represents the generation based FCR (base case) and plain lines are relative to demand-side activation (studied cases). Left: first minute. Right: first 2h.

On the other hand, frequency deviations converge to zero at a slower rate in case FCR are provided solely by loads. This is a first negative consequence of rebound errors and may be observed on the right chart of Fig. 6.5. Table 6.3 shows the frequency convergence time: the minimal time after which frequency deviations remain within the ±10mHz band.

Table 6.3: Convergence time of the frequency error to the ±10mHz dead-band.

T_n	1/4h	1/2h	1h	2h	**Base Case**
Time (s)	440	411	398	393	**387**

Longer term perspective: the rebound impacts aFRR use.
Longer term FCR activation profiles (right chart of Fig. 6.6) highlight the impact of the energy rebound. In the base case, generation units return their initial set-point smoothly and the activated FCR volumes FCR_G converge to zero. In the other cases, control errors appear. Indeed, the load group exhibits a counteracting behavior just after having provided its flexibility. The consumption of the group starts increasing (negative volumes w.r.t. FCR definition) to recover the amount of energy that has just been shifted. The error profiles vary strongly depending on the running duration T_n of the involved loads. The longer the running time T_n, the smoother the rebound and its system impact.

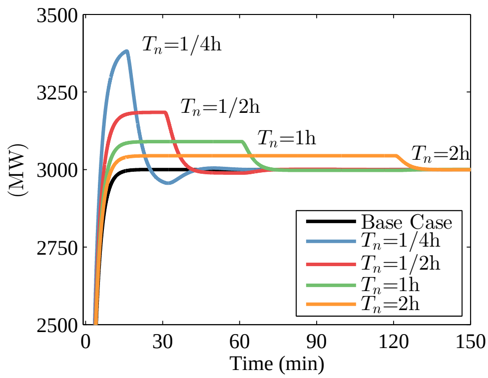


Figure 6.7: aFRR in the 150 minutes following the incident. The rebound impacts largely the use of this slower reserve capacity especially when loads with short run time participate to FCR.

Fig. 6.7 shows the aFRR activation profile following the event. Similarities between the aFRR and FCR profiles (Fig. 6.6) highlight the important influence of aFRR in the rebound error compensation. As shown, an additional amount of aFRR capacity must be contracted for rebound compensation. This negative impact may completely outweigh the initial positive impact of the loads. Table 6.4 gives some details on the required capacity increase whose costs are detailed below. Let's note that, eventually, aFRR reach the equilibrium value of 3000 MW corresponding to the initial imbalance (Fig. 6.7).

Table 6.4: Increase in aFRR capacity consecutively to the energy rebound.

T_n	1/4h	1/2h	1h	2h
FRR_{EnR}	12%	6%	3%	1.5%

Error fade-out and frequency impact.

The rebound error has a small impact on the system frequency. The major consequence appears when the error goes back to zero (error fade-out). Secondary control (aFRR) is not responsive enough to counteract the error fade-out. The resulting imbalance induces small positive frequency changes that stay within the tolerance band ±10mHz (Fig. 6.8).

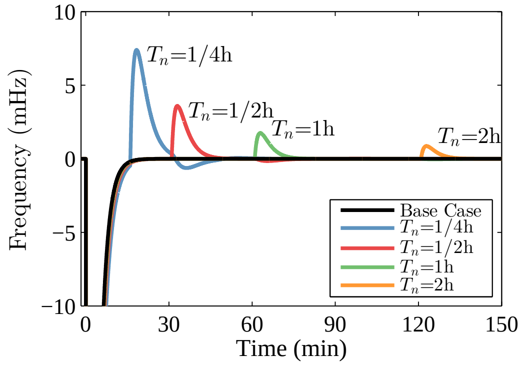


Figure 6.8: Frequency deviations during the 2h30 following the incident. Overshoots are observed as error disappears.

6.4 Historical Simulation of FCR.

In order to assess the quality of a group's response, historical frequency errors $\Delta f(t)$ and associated primary control response of the Continental Europe grid in 2013 are simulated using Eq.(6.1).

6.4.1 Simulation settings and parameters.

The system presented in equation (D.3) is simulated during one year. The aggregate models relative to the Delay and Stop/Rate control policies are used to represent the load's reaction. We explore the similar load scenarios than presented on section 5.7.

The power system is simulated with different FCR replacement levels. Indeed, the assumptions are such that the group replaces partly or totally (i.e., 10% to 100%) the 3 GW of generation assets used for *upward* FCR (negative frequency errors) in the CE synchronous area. Downward volumes (and remaining upward volumes) are provided by ideal generation represented by a first order transfer function with time constant τ_g. The droop K_{FCR} is divided into its upward and downward parts that are adapted accordingly. The related frequency deviations $\Delta f_1^-(t)$ (resp. $\Delta f_1^+(t)$) are frequency deviations limited to negative (resp. positive) values.

$$K_{FCR}^{up} + \frac{D_0 K_D}{\phi} = 15GW/Hz \qquad \& \qquad K_{FCR}^{dn} = 15GW/Hz$$

The FCR set-point is derived below.

$$FCR_G^{set}(t) = -K_{FCR}^{up}\Delta f^-(t) - K_{FCR}^{dn}\Delta f^+(t)$$

The other important system-related parameters are shown on Table 6.5.

Table 6.5: One-bus System Parameters.

Name	Value	Name	Value	Name	Value
M	25 GWs/Hz	π_M	30%	$D_0\pi_M$	[0.3-3] GW
K_{FCR}	15 GW/Hz	L	4 GW/Hz	τ_g	5 s

6.4.2 Recovering historical net imbalance.

The historical system-wide imbalance signal $I(t)$ that is at the source of frequency errors must be evaluated from historical frequency measurements. It will be used as input to the idealized system of equation (6.1) to reconstruct frequency deviations and FCR activation volumes when system parameters are modified. This allows to evaluate the overall impact of introducing e.g., demand response as reserve provider. The input reconstruction process consists in selecting an unknown input observer. Such observer is itself a system that relies on some modeling assumption. The final result will be very much dependent on the quality of the output measurements at disposal, the choice of the initial system model and the selected observer system. Several options exists, that are shortly described in appendix D.

We have at our disposal the average system frequency as measured over successive periods of 10 seconds. Very fast phenomenon and extreme frequency deviations cannot be correctly assessed from such data source. However, it is useful to assess frequency quality statistics and related flexibility activation volumes. We denote by $y_f(t)$ these frequency measurements (around $f_n = 50Hz$).

The selected power system model is denoted *immediate damping with integral action for rebound management*. Shortly, the imbalance $I(t)$ is recovered by applying a constant proportional factor to historical frequency measurements. Then, a layer of aFRR is considered in the model that will only be used for counteracting the rebound action. We cannot reconstruct more realistic imbalance $I(t)$ from the available data. In practical situation, aFRR basically counteract the integral of the frequency deviations $\Delta f(t)$. However, modeling aFRR in such way would require to include, in the input recovery process, the reaction of RR and other balancing resources as well as the non-ideal nature of both aFRR and FCR (saturation, dead-band, decentralized aspect of aFRR in LFC area, etc.). Indeed, the integral of the frequency measurements we have at disposal is quite high. If pure aFRR had to counteract this integral, they would reach unrealistic levels of deployment. In order to obtain realistic aFRR deployment levels in simulations, and explain the relatively high frequency integral, the non-ideal elements should be taken into account. This is discussed on appendix D in more details.

6.4.3 Simulation model

The model being simulated is shown below for clarity. It is slightly different from the one-bus system model. In practice, we use a discrete time version of the model below.

$$M\frac{d\Delta f(t)}{dt} = I_1(t) - L\Delta f(t) + FCR_G(t) + aFRR(t) - D_0\hat{x}^j(t) \quad (6.3)$$

$$\tau_g\frac{dFCR_G(t)}{dt} = (-K_{FCR}^{dn}\Delta f^+(t) - K_{FCR}^{up}\Delta f^-(t)) - FCR_G(t) \quad (6.4)$$

$$\tau_{g,2}\frac{daFRR(t)}{dt} = aFRR^{set}(t) - aFRR(t) \quad (6.5)$$

$$aFRR^{set}(t) = -K_{aFRR}\left(ACE_G(t) + \frac{1}{\tau_{aFRR}}\int ACE_G(t)dt\right) \quad (6.6)$$

$\Delta f(t)$	Hz	Frequency deviation at time t.
$I_1(t)$	MW	System power imbalance at time t (immediate damping).
M	MWs/Hz	Angular Momentum of the rotating masses.
L	MW/Hz	Natural frequency-dependence of total system load.
$FCR_G(t)$	MW	FCR volumes provided by generation.
$aFRR(t)$	MW	aFRR volumes provided by generation.
$\hat{x}^j(t)$	MW	Group relative flexibility (heterogeneous aggregate model).
D_0	MW	Initial demand of the group.
τ_g	s	Time constant of FCR generation resources.
K_{FCR}^{dn}	MW/Hz	Downward FCR droop for generation assets.
K_{FCR}^{up}	MW/Hz	Upward FCR droop for generation assets.
$aFRR^{set}(t)$	MW	set-point of aFRR resources.
		$\Rightarrow$ PI control with gain K_{aFRR} and time constant τ_{aFRR}.
$ACE_G(t)$	MW	Area control error sent to aFRR resources.
$\tau_{g,2}$	s	Time constant of aFRR generation resources.

As can be seen, aFRR resources react with a certain time constant to the set-point delivered by a proportional-integral controller (PI) which take error $ACE_G(t)$ as input.

Defining the ACE (integral action)

The challenge is to find the most representative control error $ACE_G(t)$ that will be sent to aFRR resources. Indeed, we would like to represent the situation presented in the event-based simulation case where aFRR resources participates to counteracting the rebound error. This error is not measurable in practice but appears indirectly in the frequency deviation. Yet, as discussed above, aFRR are cannot be directly linked to the frequency deviation in the historical simulation context.

For illustration purpose, we compare two situations. The first is an ideal situation where the rebound error is perfectly known and sent to aFRR resources. In such case, the frequency quality is degraded if the group cannot provide the correct damping or if its rebound error is not fully compensated (e.g., ramp rate limit of aFRR, etc.).

$$ACE_G^{ideal}(t) = -D_0(\hat{x}^j(t) - K_D r^{up}(t)) \tag{6.7}$$

The second option is to use aFRR resources to counteract all frequency changes differing from the historical ones. This means that aFRR will both counteract the rebound but also take the role of FCR (though consistently slower) if the group does not provide sufficient damping (e.g., Delay policy). The resulting frequency quality will be close to the historical one except for its fast-varying part. Indeed, the slow ramp rate of the aFRR resources prevents them to deliver the same damping than what is expected from FCR. As loads are restricted to provide upward volumes, the ACE accounts for the negative frequency errors obtained in the simulation $\Delta f^-(t)$ that are larger (i.e., more negative) than their historical counterpart $y_f(t)^-$.

$$ACE_G^{historical}(t) = K_{FCR} \left. \frac{\Delta f^-(t) - y_f^-(t)}{\phi} \right|_{<0} \tag{6.8}$$

The two ACE definitions lead to different results , that we highlight in a one day simulation on figure 6.9. This figure highlights the frequency quality before (initial) and after the introduction of ECL control. The ECLs are supposed to cover the full upward FCR range following either the Delay or the Stop/Rate control policy. Each chart is relative to one of the above ACE definition.

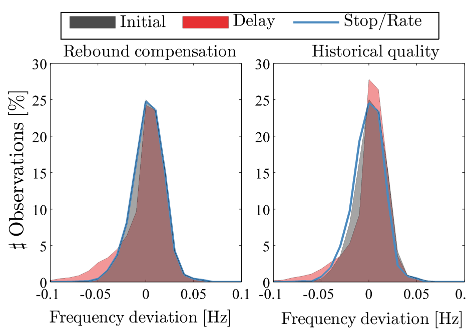


Figure 6.9: General frequency behavior for the two ACE definitions. Left chart: ideal rebound compensation. Right chart: historical quality target.

As can be observed in the first chart of figure 6.9, the results obtained using the Stop/Rate policy are almost similar to the initial data. The Delay control policy is effective in counteracting small frequency changes. As soon as frequency variations are larger, the limited ramp rate start being constraining. This explains why larger negative frequency changes are observed when loads are controlled under the Delay policy. On the right-hand chart, the second ACE definition leads aFRR to overreact. Indeed, frequency is pushed upward by the aFRR reaction.

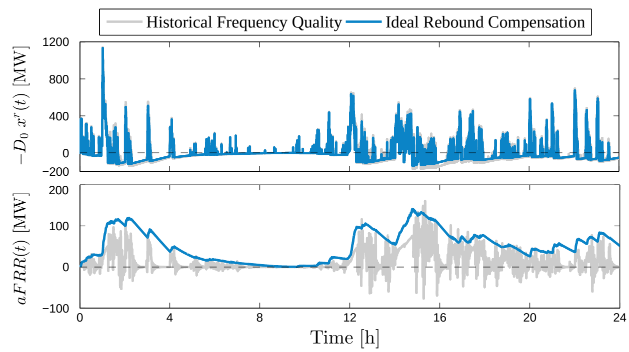


Figure 6.10: aFRR activation under the two ACE definition. Stop/Rate policy.

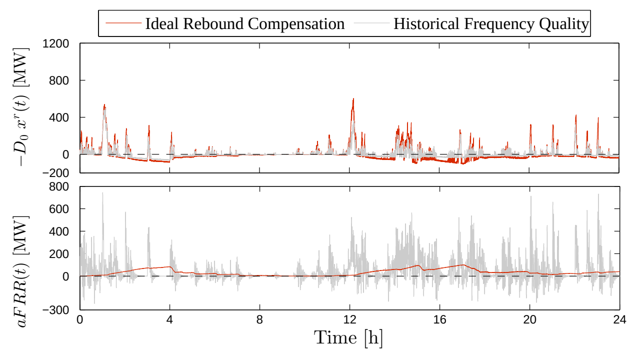


Figure 6.11: aFRR activation under the two ACE definition. Delay policy.

The most realistic case seems to be the ideal rebound compensation. This is particularly visible from the two following figures (Fig. 6.10 and 6.11) that show the aFRR activation volumes in the two considered cases. The aFRR activation profiles obtained through simulations look much alike their realistic behavior in the ideal compensation case.

A zoom into frequency dynamics.

For illustration purpose, we show on figure 6.12 the frequency changes as simulated from the first quarter of hour of 2015, where loads overtake the complete upward FCR volumes. The initial frequency data are shown on the top chart of figure 6.12 (stair shape), and compared to the simulated frequency change (dash-dot) obtained via the recovered input $I_1(t)$. The two other charts represent respectively the delay and Stop/Rate control policy. In both of these charts, the dashed lines are relative to scenario I and plain lines to scenario II. As a recall, scenario I accounts for loads with small run time (i.e, $T_n \in [1/4h, 1/2h]$) and scenario II for loads with larger energy needs (i.e., $T_n \in [1/4h, 3h]$). In what follows, the term *with DR* (demand response) refers to cases where loads are actively controlled.

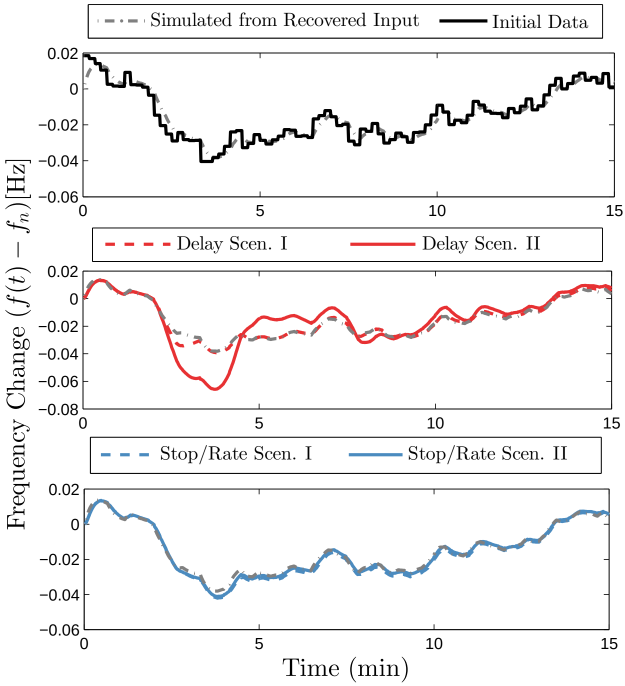


Figure 6.12: Frequency deviation from initial data, simulated from recovered input first without and with DR under the two control policies (DR covers the full upward FCR).

In the Stop/Rate control policy, the frequency evolves similarly with and without DR. However, in the Delay policy, only loads with short run times (scenario I) lead to acceptable performances. In this control policy, the ramp rate of the group demand change is inversely proportional to the run time of the involved loads (see $x^d(t|T_n)$ in (C.1)). However, the gain K_D is set to 50% according to the the assumed load technical capability but does not take into account the group ramp rate limit. As can be observed on figure 6.12, loads with short run times react relatively well for small frequency deviations. These good technical performances start degrading for larger and fast varying frequency errors.

These performances are assessed rigorously below.

6.4.4 Response quality with ideal ACE definition

We show below the response of the group $\hat{x}(t)$ as a function of the normalized frequency deviation $r(t)$ in both simulated scenarios (1-year simulation). As the number time steps in the simulation is very large, we represent the simulation output in the form of convex hulls. Let's define the simulation output couple ($\hat{x}(t)$,$\Delta f(t)/\phi$) for each time t in the simulation. The associated set that gathers the result of the full simulation is ($\mathcal{X}$,$\mathcal{F}$). Let's distinguish two subset $(\mathcal{X}^-, \mathcal{F}^-)$ and $(\mathcal{X}^+, \mathcal{F}^+)$ that are respectively relative to outputs where $\Delta f(t) < 0$ and $\Delta f(t) \geq 0$. We compute the union of the convex hulls of each subset, which results in a set denoted $(\mathcal{X}_{hull}, \mathcal{F}_{hull})$.

$$(\mathcal{X}_{hull}, \mathcal{F}_{hull}) = Conv(\mathcal{X}^-, \mathcal{F}^-) \cup Conv(\mathcal{X}^+, \mathcal{F}^+) \tag{6.9}$$

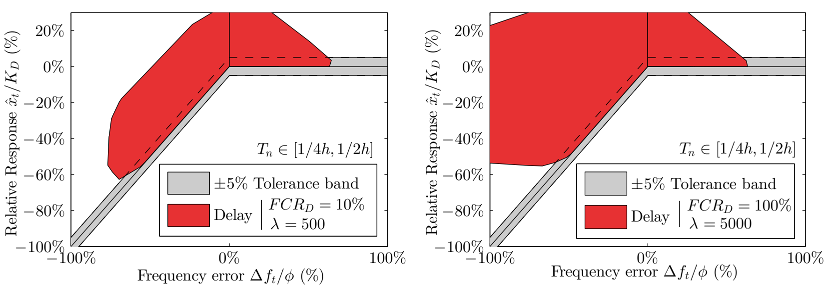


Figure 6.13: Delay performance for 10% and 100% FCR replacement level (Scenario I).

In scenario I, the performances of the Delay control policy are poor (Fig. 6.13), even though quite good for small frequency changes ($\Delta f(t)/\phi >$ -30%). The red sets indeed jumps rapidly out of the tolerance band. The performance of the Stop/Rate policy are unsatisfactory as well (Fig. 6.14).

In Scenario II, the performances of the Delay control policy are even poorer than in scenario I (Fig. 6.15). In the Delay policy, the dominant

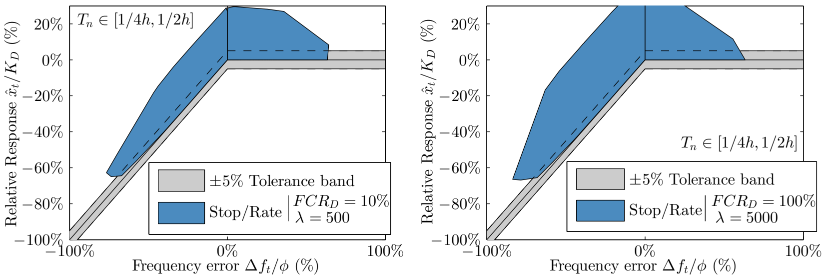


Figure 6.14: Stop/Rate performance for 10% and 100% FCR replacement level (Scenario I).

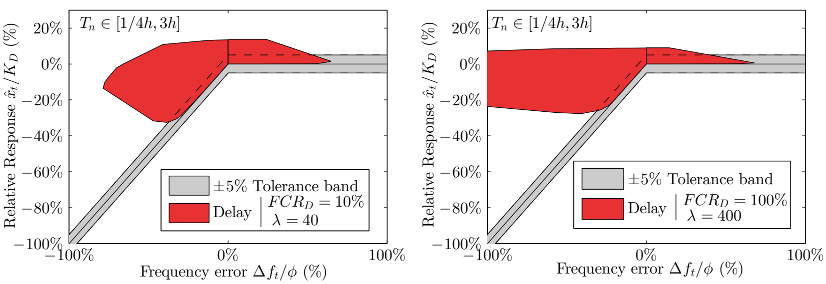


Figure 6.15: Delay performance for 10% and 100% FCR replacement level (Scenario II).

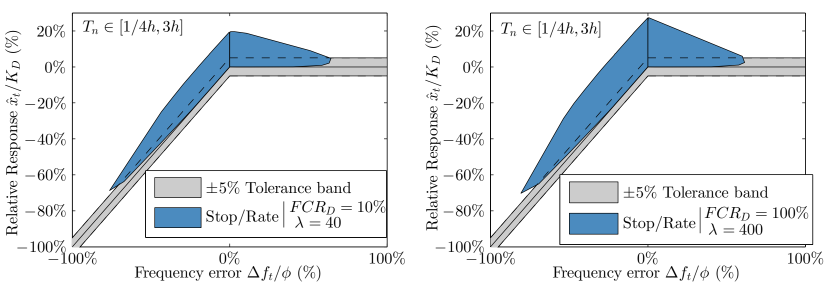


Figure 6.16: Stop/Rate performance for 10% and 100% FCR replacement level (Scenario II).

source of error is the limited ramp capability of the group. The provided flexibility cannot ramp down faster than D_0/T_n (homogeneous case). The delay strategy behaves therefore better with loads of small duration (Scen I). On the other hand, the performance of the Stop/Rate are very good

(Fig. 6.16), and detailed below.

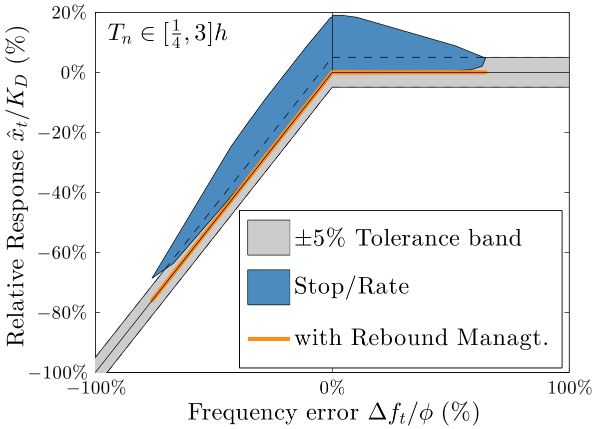


Figure 6.17: Convex Hull of the group's Response w.r.t. frequency deviations in the second Scenario. The net response (with rebound management from aFRR) is also shown.

As can be observed, the rebound management is very effective. The most important feature is that the answer gets more precise as the deviation are larger. This is a desirable feature as the group responds perfectly well in case the system is subject to large disturbance. The response of the group together with the aFRR contribution is providing an almost perfect response. If the aFRR contribution is omitted, the group could be perceived as providing an inadequate service. Indeed, the operator would notice that a certain number of points are out of the tolerance band.

The only way to overcome this issue requires a threefold action (1) slight increase of the tolerance band to ±10%, (2) select loads with larger run time and (3) virtually decrease the baseline to benefit from the negative tolerance band. The effect of these actions are illustrated on figure 6.18.

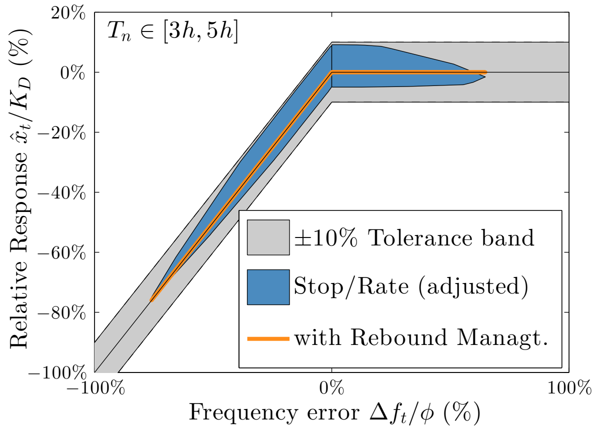


Figure 6.18: Adjusted Convex Hull of the group's Response w.r.t. frequency deviations for loads with long run time (here : $T_n \in [3, 5]h$).

6.4.5 Consequences of Rebound management

In the long-run, energy rebounds may have a strong impact on the system frequency distribution. To counteract this trend, an additional layer of automatic Frequency Restoration Reserve (aFRR) *slowly* compensates for energy rebound. The total amount of extra FRR capacity FRR_{EnR} that were required in the simulation for this specific purpose is shown on figure 6.19.

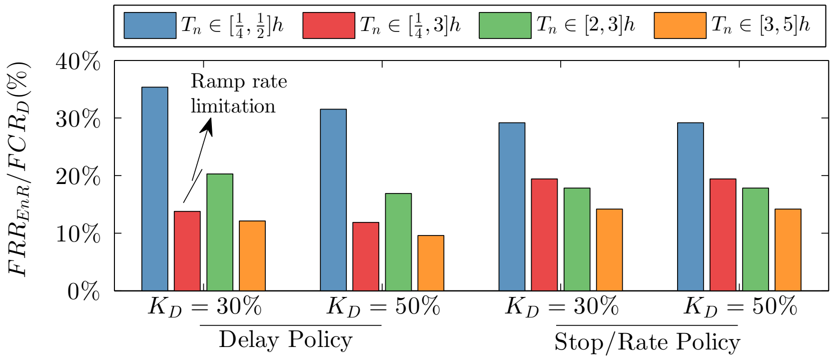


Figure 6.19: Effect of Run Time and Droop K_D on the additional aFRR capacity $aFRR^+$ required for rebound compensation.

The required capacity FRR_{EnR} (MW) depends on run time of the involved loads and on the selected Droop K_D. It is shown on figure 6.19 in relative proportion of $FCR_D = 3GW$, the FCR volumes covered by the flexible demand. Results are shown for each control policy and for different run time distributions. Loads with longer run time induce lower extra aFRR needs. Yet, the Delay policy is unable to maintain historical frequency quality level (see Figure 6.9). A higher droop K_D tends to decrease the aFRR needs in the Delay policy (see $K_D = 30\%$ and $K_D = 50\%$ on 6.19). The ramp rate constraint is indeed more binding for larger K_D, which in turn, degrades performances. As this simulation exploits the ideal ACE definition, aFRR volumes do not perceive this degradation. They are simply less solicited.

Performances of the Stop/Rate policy exploiting loads with larger run time are good. A positive business case could emerge from this policy. This is discussed in details below. The bar heights of the third (or fourth) chart in figure 6.19 are compared to the event-based results in Table 6.6. Historical simulations lead to higher FRR needs than what was obtained in the event-based simulation.

Table 6.6: Increase in aFRR capacity due to energy rebound.

Event-Based - 100% DR - Stop/Rate Policy				
T_n	$\frac{1}{4}$h	$\frac{1}{2}$h	1h	2h
FRR^{eb}_{EnR}	12%	6%	3%	1.5%
Historical - 100% DR - Stop/Rate Policy				
T_n	$\frac{1}{4}$h-$\frac{1}{2}$h	$\frac{1}{4}$h-3h	2h-3h	3h-5h
FRR^{hist}_{EnR}	30%	21%	17%	14%

Practically speaking, the system operator will need to contract an additional amount of capacity FRR_{EnR} for rebound compensation. The sourcing of such extra capacity can be costly. The overall economic impact depends on the relative costs of FCR and FRR capacity. A detailed cost-benefit analysis is conducted in the following section.

6.4.6 Rebound management with the historical quality ACE definition.

Stop/Rate policy results

The group's response exhibit lower performance with the second ACE definition (figure 6.20).

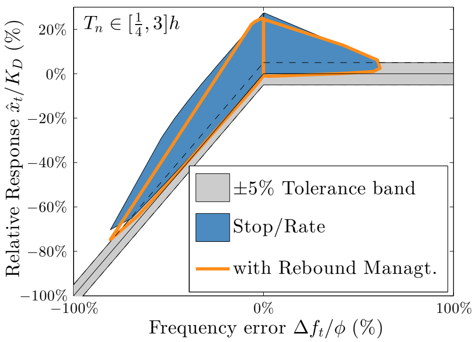


Figure 6.20: Convex Hull of the group response (Stop/Rate, 2^{nd} ACE def.).

In the Stop/Rate policy, the rebound management is indeed less effective but the frequency quality can actually be maintained very close to its initial level.

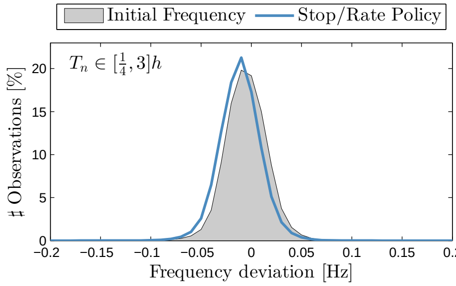


Figure 6.21: Distribution of the System Frequency deviation: initial vs simulated case (Stop/Rate policy). Rebound compensation via the frequency error.

As can be seen on figure 6.21, the frequency tends to be a little lower in this case. The information must now transit via the frequency signal. The frequency will tend to go down due to the rebound error but will be pushed back upward by the slower aFRR resources.

Delay policy results

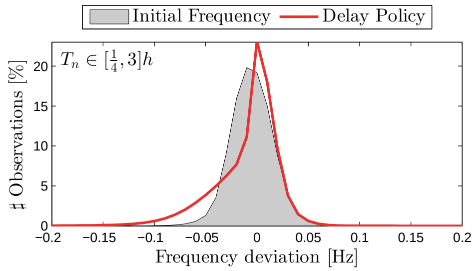


Figure 6.22: Distribution of the System Frequency deviation : initial v.s. simulated case in the Delay policy. Rebound compensation via the frequency error.

In the Delay Policy, the limited ramp rate of the demand-side response cannot be fully compensated by the slower FRR volumes. The frequency quality is much worse (Fig.6.22) while the extra capacity required is very high (Fig.6.23).

Rebound management and FRR capacity increase

The aFRR needs a very high in the Delay policy (Fig. 6.23). In the Rate/Stop case, the aFRR capacity is lower with this second ACE definition. This means that frequency quality can be maintained closely to historical results, even though the rebound is not fully compensated.

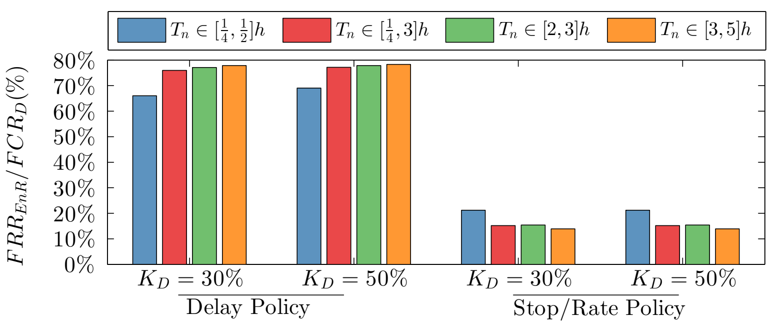


Figure 6.23: Additional aFRR capacity FRR_{EnR} required for rebound compensation. The rebound is compensated via the frequency error.

6.5 Costs and benefits at system level

We assess in this section the economics of the Stop/Rate control policy. The cost-effectiveness of the proposed solution strongly depends on the local system context.

6.5.1 System Benefits from introducing load control

The generated yearly profits π_{FCR} come from avoided payments to traditional FCR resources, expressed below in millions of euro. A factor γ is 2 in the symmetric reserve case, and 1 in case loads are restricted to provide an asymmetric service.

$$\pi_{FCR} = \gamma\, FCR_{cap}\, \bar{p}^{FCR}\, 8760 \tag{6.10}$$

The FCR capacity price $p^{FCR}(t)$ and its average value $\bar{p}^{FCR}$ in the Belgian system are shown on figure 6.24. Each bar represents the asymmetric[2] price of each 2015 monthly tendering period and is relative to a specific product of the Belgian system.

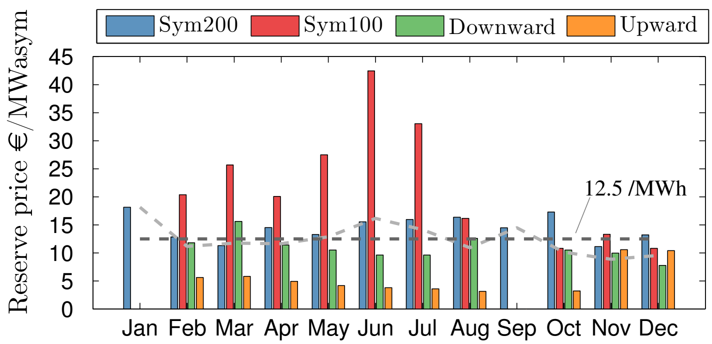


Figure 6.24: FCR monthly prices by product type (Belgium, 2015) [47].

In the Belgian system, FCR are divided into four different products : symmetric 200, symmetric 100, upward and downward. They are briefly presented in the power frequency curve of Figure 6.25. There are two symmetric and two asymmetric products.

6.5.2 Additional System Costs

The ECL response also involve costs of which rebound management and demand-side costs (controller) are the main source.

[2]Must be counted twice in the revenues of a symmetric resource.

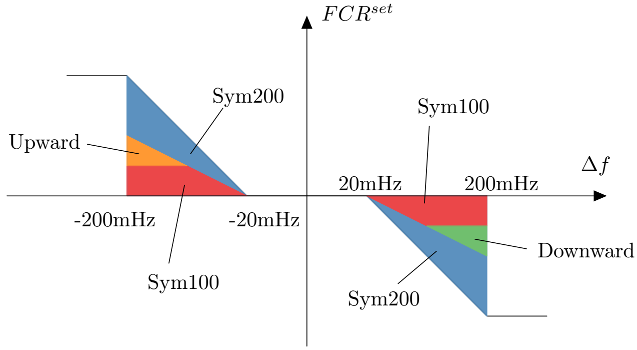


Figure 6.25: The FCR products, Belgian system 2015. Source: Elia [47].

Cost of capacity: FCR vs aFRR

A first part of yearly additional costs C_{aFRR} originates from extra aFRR capacity $FRR_{EnR}(T_n)$ required to cover the rebound. It is dependent on the load's run time T_n (or scenario). The price for $\bar{p}^{aFRR}$ is chosen as its average in the Belgian market in 2015 (Figure 6.26).

$$C_{aFRR} = \gamma\, FRR_{EnR}(T_n)\, \bar{p}^{aFRR}\, 8760 \tag{6.11}$$

Due to the rather small energy content of FCR, costs related to extra activation payment of aFRR are neglected. Consequences on capacity reservation costs are indeed much higher. We will consider both event-based and historical simulation cases to account for the required FRR_{EnR}.

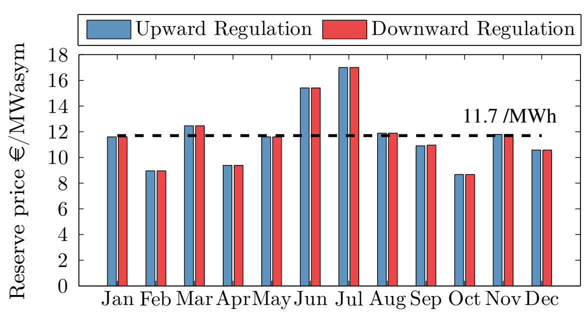


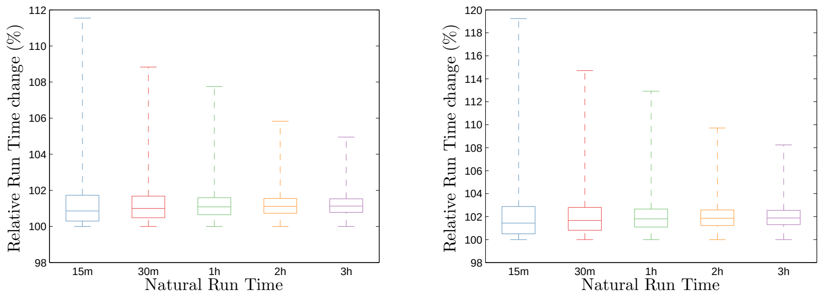
Figure 6.26: aFRR monthly prices in the Belgian area (2015). Source: Elia [47].

Demand-side Costs

Even when externally controlled, loads behave very much as scheduled in their initial planning. Indeed, the energy content of the service they provide is small enough such that consequences on run time are small. FCR capacity is designed to cover extremely rare and very large events. Most of the time, the percentage of available capacity that is actually deployed is quite small. The consequences on loads run time can be estimated from the frequency signal.

$$\frac{T^{run}(t, T_n)}{T_n} \simeq 1 + \frac{K_D}{\phi} \frac{1}{T_n} \int_0^{T_n} \Delta f^-(t-s)\, ds \tag{6.12}$$

Results of the approximation (6.12) are illustrated for loads with different natural run time in figure 6.27a and 6.27b respectively for $K_D = 30\%$ and $K_D = 50\%$. Several box-plot charts highlight the distribution of the relative run time increase that would have been experienced by loads exploited for upward FCR in 2015[3].



(a) $K_D = 30\%$ (b) $K_D = 50\%$

Figure 6.27: Run Time change distribution due to FCR participation.

As can be seen on both of the above figures, the largest run time change occur rarely. Users should not get compensated for such small change in consumption duration. Therefore, demand-side costs are limited to controllers' installation costs.

6.5.3 Cost-benefit analysis: maximum controller cost

A cost-benefit analysis of the above results is conducted. The main assumptions are presented below. Our goal is to find the maximum allowed cost

[3]Results from a 1 year simulation, from Nov.2014 to Oct.2015.

for the power controller that would lead the net present value NPV to be zero after a 5 year period. Considering an initial cost (installation costs) of C_0 and successive net revenues $NR(k)$ (i.e., net cash flow) for each year $k \in [1, N_y]$, the net present value of the program after N_y years is defined below. We apply a discount rate r_d (cost of capital) on each year's revenue.

$$NPV(N_y) = -C_0 + \sum_{k=1}^{N_y} \frac{NR(k)}{(1+r_d)^k} \tag{6.13}$$

The cost of capital r_d is set to 10%, which corresponds to a medium level of risk according to [54]. In case of mandatory participation programs (e.g., standards), this is an overestimate. The goal is therefore to derive the maximum total cost $C_0(5)$ such that the NPV is zero after 5 years.

$$C_M^{5y} = \sum_{k=1}^{5} \frac{NR(k)}{(1+0.1)^k} \tag{6.14}$$

The yearly revenues are constant $NR = \pi_{FCR} - C_{aFRR}$. The price for FCR and aFRR are conservatively set to respectively 10/€/MW and 12/€/MW. These prices are relative to a one-directional service.

The maximum allowable cost per participating load is $c_M^{5y} = C_M^{5y}/n_l$. Results are shown on figure 6.28 for $P_n = 2kW, K_D = 50\%$ loads with different parameters and products (asymmetric upward or symmetric flexibility).

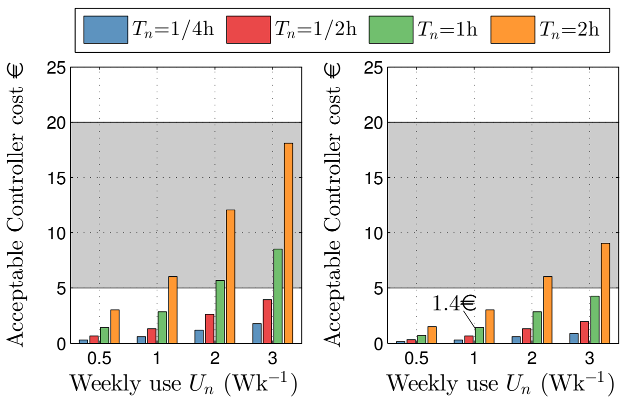


Figure 6.28: Maximum acceptable cost of the power controller as a function of the load's run time and weekly use. Parameters: $P_n = 2kW, K_D = 50\%$. Left chart: symmetric reserve. Right Chart: assymetric upward reserve.

We consider loads with different weekly utilization rate U_n which is the average number of time that a load is requested to start on a weekly basis.

As can be seen on figure 6.28, a filled grey area indicates a range of possible controller costs estimated to lie between 5€[57] and 20€[66]. Let's note that, with these simple controllers, the proportional gain K_D/ϕ cannot be remotely changed.

The control makes economic sense for loads providing a significant *potential flexibility* F_{wk} per week. The *potential flexibility* per week $F_{wk} = K_D U_n E_n$ is defined as the product of the appliance average energy consumption per week ($= U_n E_n$) with the maximal relative change that the appliance can impose to its power rate ($= K_D$). Let's note the large influence of U_n, which increases per load revenues, as benefits must be divided among a lower number of participants.

For simplicity, we have assumed an ideal case where loads start regularly all along the day, week, month and year. The required number n_l of participating loads to replace the traditional FCR capacity (i.e., 3000 [MW]) is inversely proportional to the potential flexibility F_{wk} offered by each load. It is indeed proportional to the starting rate λ required for loads consuming an energy E_n to have a sufficiently large baseline power D_0 such that $K_D D_0 = 3000MW$. When λ is expressed in number of start per hour (i.e., or E_n in kWh), the formula below can be used to assess the number n_l of participating loads.

$$n_l = 24[\mathrm{h.d}^{-1}]\,\lambda[\mathrm{h}^{-1}]\frac{7\,[\mathrm{d.wk}^{-1}]}{U_n\,[\mathrm{wk}^{-1}]} = \frac{3.10^6}{K_D E_n}\frac{7\times 24}{U_n} = \frac{504}{F_{wk}} \quad [mio] \tag{6.15}$$

On Table 6.7, detailed results are presented for 2kW loads delivering symmetric capacity (i.e., $\gamma = 2$) reaching up to 50% of their initial power and running on average twice a week (i.e., $U_n = 2$). The largest influence on overall profits comes from the required number of active loads n_l (details below). The extra aFRR reservation costs are of lower importance. Let's note that, the FCR/FRR price spread has been chosen conservatively.

Rebound error compensation within balancing markets

It is possible to improve the above results. In fact, the ECL group and aFRR are the only resources capable of providing sufficient power/energy for covering the N-2 criterion (event-based simulation). Therefore, the required capacity FRR^{eb}_{EnR} resulting from the event-based is a lower bound for the capacity increase.

However, in more normal situations, covered in the historical simulation, the slowly evolving component of the frequency error could be partly compensated by other resources than aFRR. The capacity increase FRR^{hist}_{EnR}

Table 6.7: CBA results ($\gamma = 2$, $P_n = 2$, $K_D = 50\%$, $U_n = 2$).

Load Duration T_n	h	1/4	1/2	1	2
Load Flexibility F_{wk}	kWh/wk	0.5	1	2	4
Starting rate λ	s^{-1}	3340	1670	835	413
Required number of loads n_l	(mio)	1000	500	250	125
Annual energy	kWh/load	50	100	200	400
Slow reserve needs FRR_{EnR}^{hist}	MW	1000	850	700	600
aFRR yearly costs C_{aFRR}	(mio €/y)	210	180	150	130
FCR yearly benefits π_{FCR}	(mio €/y)	525	525	525	525
DR yearly profits	(mio €/y)	315	350	375	400
Per Load yearly profits	(€/load /y)	0.31	0.7	1.5	3.2
Controller's cost	**€**	**1.2**	**2.6**	**5.7**	**12**

obtained in the historical simulation is therefore an upper bound. In practice, RR or even the balancing market (non-reserved) could cover the most part of the difference between the upper and lower bound of FRR_{EnR}.

$$FRR_{EnR}^{eb} \leq FRR_{EnR} \leq FRR_{EnR}^{hist} \tag{6.16}$$

Indeed, as the energy component of the frequency deviations is slowly evolving the system operator would have time to call for less expensive capacity in order to relieve aFRR from covering the rebound error. The consecutive costs that would be required to pay for the flexible energy volumes in the balancing market are negligible with regards to the reservation costs of aFRR capacity [88].

The results of the cost-benefit analysis when FRR_{EnR} are considered at their lower bound (event-based simulation) is shown on figure 6.29.

In this case, loads with a run time around 30 minutes become possibly interesting. Yet, the results to not change fundamentally. This suggests that the main driver of the maximum controller cost is the number of involved loads, and therefore the annual energy consumption of the load. This can be e.g., visualized from comparing the bar ($U_n = 0.5, T_n = 2h$) with bar ($U_n = 1, T_n = 1h$) that have almost the same height, as they are relative to loads with the same annual energy consumption. We won't enter into longer discussions, but the interested reader can find some details about perfect forecast energy rebound compensation within balancing markets in [88].

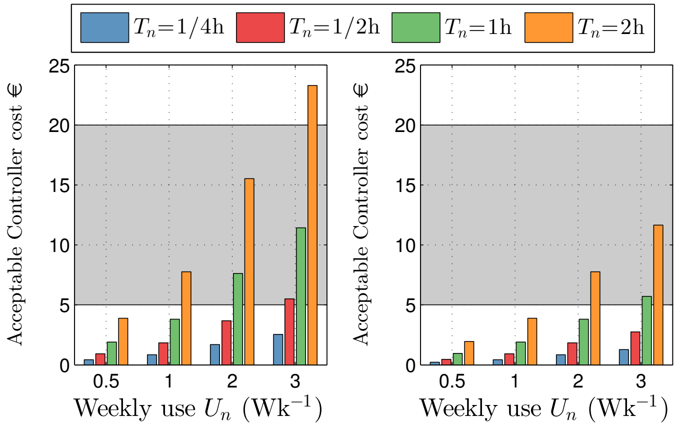


Figure 6.29: Maximum acceptable cost of the power controller when extra aFRR needs are set to their lower bounds (event-based simulation). Parameters: $P_n = 2kW, K_D = 50\%$. Left chart: symmetric reserve. Right Chart: assymetric upward reserve.

Doubling the return period

The same results are shown on figure 6.30 when considering that the NPV should be zero after 10 years, instead of 5, and that the aFRR needs are at their upper bound.

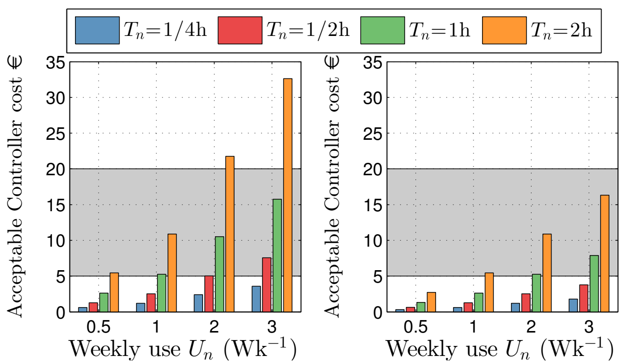


Figure 6.30: Maximum acceptable cost of the power controller for $NPV = 0$ after 10 years. Parameters: $P_n = 2kW, K_D = 50\%$. Left chart: symmetric reserve. Right Chart: assymetric upward reserve.

6.5.4 Conclusion

Installing simple controllers seems beneficial on the overall system perspective for loads consuming at least 100 − 200kWh annually, most probably even 300kWh, and running for a minimum of 30 consecutive minutes. This annual energy consumption can be reached by batch water boilers or small industrial pumps and are way below electric vehicle's consumption. In the more conservative case, loads consuming about $100kWh$[4] annually require ultra low cost controllers of about 1.4€for loads when restricted to deliver asymmetric flexibility.

[4]E.g. small dishwashers or refrigerators, though this model does not capture the dynamics of TCL.

Chapter 7

Power control with rebound management

Chapter summary

Rien ne sert de courir, il faut partir à point (La Fontaine, 1668). As the tortoise's hard work surpasses the hare's idleness, can a load group parade in a more convenient carapace ? What would it cost for loads to counteract their nature and store energy a little longer ?

Energy rebounds are source of extra costs as they call for additional slower reserve capacity (aFRR) which are expensive. This chapter discusses an advanced control framework which essentially allows to store energy temporarily within the load group instead of counting on the slower reserves for rebound compensation. Consecutively, the additional efforts undertaken on the group-side require part of the initial flexibility to be excluded from FCR which in turn decreases the per load profitability of this control scheme.

7.1 Objectives

The objective of this chapter is to design an autonomous power controller able to deliver FCR while compensating for the rebound error. This supposes that some energy management elements are integrated. Indeed, as was shown in the previous chapters, any attempt to control a group's power always leads to some energy to be stored within the group. The amount of energy that can be stored is limited by the load's characteristics and by the end-user preferences. Therefore, some energy recovery mechanism must be integrated into the local controller. Based solely on local frequency measurements, on the parameters of the load being controlled and on this energy recovery mechanism, this controller must be able to assess how much firm flexible capacity it can deliver (capacity allocation) to the TSO. In addition, it will need to compute the appropriate load power set-point autonomously.

In this chapter, we assume that the load's power can be controlled continuously within certain limits $P(t) \in [P_L, P_H]$ (rate strategy). In a first part, we study homogeneous groups. The power $P(t)$ of each load is a unique control input that applies to all running loads in the group. Assuming a constant arrival rate, we compute the feasibility set gathering all reference signals (power) that can be tracked by the group's demand with full rebound error compensation. Such signals have bounded power and energy content. The power and energy limits depend on the load and user-related parameters. Due to these limits, an energy management process must take place, limiting the energy content of the reference signal that the group demand must track. The consequences are that the group stops following perfectly the frequency of the network. It induces therefore an other rebound error, but whose power profile can be decided in advance, and that can be predicted by the power system operator. The operator can therefore decide to compensate the forecast rebound error in an optimal way.

In a second part, we extend the analysis to heterogeneous groups. As we do not consider any communication infrastructure, the loads continue to behave as if they were part of an ideal homogeneous group (as was done above). The single consequence of the heterogeneous nature of the group happens on the system operator's side: it must assess, given the parameter distribution and the energy recovery mechanism at stake, the amount of FCR that is being provided by the running loads, and the consecutive rebound error that needs to be compensated.

7.2 Inverse modeling power control in Homogeneous group.

Let's take a look at all constitutive elements of the advanced power controller illustrated on figure 7.1 : (1) energy management (E.Mgt), (2) capacity allocation and (3) inverse model. An indirect feedback exists between the group's demand and the frequency deviations.

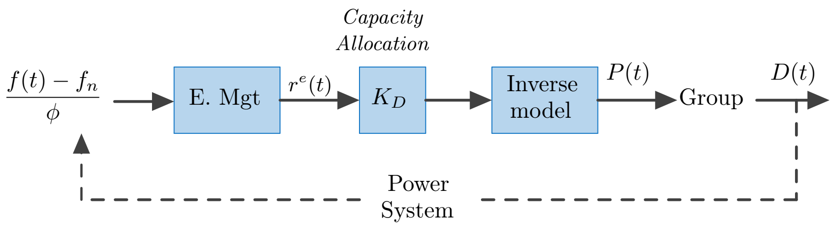


Figure 7.1: Overview of the constitutive elements of the closed-loop power controller able to counteract rebound errors for a limited set of reference signals.

7.2.1 Energy management

The energy management block limits the energy content of the reference signal that the group must track. Practically, it can take multiple forms.

1. High-pass filtering of the frequency measurements for offset removal. Loads counteract the fast-varying component of the measured frequency deviations while slower component (energy content) is compensated by slower resources. This is already implemented today for aFRR-like products (e.g., DReg-AReg signals in the PJM system [120]) and has been extensively discussed in the literature (e.g., [107, 23]).

2. Autonomous coordination process. Loads autonomously return to a predetermined state on a regular basis [159].

3. Ex-post correction through the use of limited communication (decentralized control). The loads react autonomously in the short-run. At the same time, the system operator computes an energy recovery trajectory that is communicated to each load on a regular basis.

The purpose of energy management is actually to transform the shape of the natural rebound error into a more desirable power profile and spread the energy recovery on longer time periods. The recovered energy is provided by slower reserve (i.e., aFRR and/or RR) and by the system as a whole (e.g., frequency deviation, natural damping). Delaying the natural rate at which energy is recovered (rebound error) requires some energy storage capability. A portion of the storage capacity of the group must be specifically reserved and used for this purpose alone. The power that can be delivered as FCR gets thereby limited. Let us also note that the energy management process has another advantage when implemented locally: it prevents error propagation [88].

7.2.2 Capacity allocation

Assuming that an energy management process is implemented, the group's demand must now respond to energy bounded frequency deviations $\Delta f^e(t)$ defining the reference signal $r^e(t) = \Delta f^e(t)/\phi$. The capacity allocation is a feasibility problem which aim is to answer the following question.

Given the energy bounds of the filtered frequency deviation and the load parameters:

$$\text{Bounds: } |\textstyle\int \Delta f^e(t)dt| \leq c$$
$$\text{Parameters: } [P_n,\, T_n,\, E_n,\, P_L,\, P_H,\, T_{dl}]$$

What is the highest gain K_D (droop), or part of the group's demand ($= K_D D_0$), that can be offered as flexible capacity and guarantees perfect tracking performances ?

7.2.3 Inverse model

An inverse modeling control framework is exploited to define the power set-point $P(t)$ of all running loads. Given the droop K_D and the input signal $\Delta f^e(t)$, the group must track a bounded reference signal $K_D r^e(t)$ with perfect precision.

The equation system that must be inverted is the aggregate group demand $D(t)$. For convenience, we denote the dynamic run time $T^{dyn}(t, T_n)$ by $\tau(t)$. We use the notation defined in [83] to express time delayed variables: $x_t(\tau) = x(t - \tau)$.

$$\mathcal{S} = \begin{cases} \dfrac{d\tau(t)}{dt} = 1 - \dfrac{P(t)}{P_t(\tau(t))} & (7.1) \\ D(t) = \lambda\, \tau(t)\, P(t) & (7.2) \end{cases}$$

with : $\tau(0) = \tau_0$; $P(t) \in [P_L, P_H]\, \forall t$; $0 < P_L < P_n < P_H < \infty$; $P(t \leq 0) = P_n \nu(t)$ (known).

Let's note that, with known limits $0 < P_L < P_H < \infty$, a natural rate P_n can always be defined to have $0 < P_L < P_n < P_H < \infty$. This means that the group may deliver upward and downward flexibility.

Equation 7.1 is the time derivative of the energy conservation constraint (recalled below in eq.(7.3)). It defines implicitly the time-varying delay (dynamic run time) $\tau(t)$.

$$\int_0^{\tau(t)} P_t(s)\, ds = E_n \quad \forall t \tag{7.3}$$

Supposing an energy bound c of the reference signal $r^e(t)$, and the consecutive symmetric droop $K_D(c, T_n, p_L, p_H)$, the inverse problem below must be solved by each load to define its normalized power set-point $p(t) = P(t)/P_n$ (i.e., the above system is normalized).

$$\frac{d\tau(t)}{dt} = 1 - \frac{p(t)}{p_t(\tau(t))} \tag{7.4}$$

$$p(t) \quad = K_D \frac{r^e(t)}{\tau(t)} \tag{7.5}$$

This system must be solved autonomously by each load. Let's imagine a load arriving at $t = 0$. Solving the above system will require the knowledge of both $\tau(t = 0)$ and all previous inputs $p(t), \forall t \in [-\tau(0), 0)$ which themselves dependent on all values in the interval $[\tau(-\tau(0)), -\tau(0))$, etc. Fortunately, this system has an interesting property: its evolution becomes asymptotically independent on the exact *power profile* that was tracked in the past. What counts is the *energy* of the past reference values. We must therefore design an energy management process that would allow the loads to compute this energy value. This will be performed by observing the reference signal for a sufficiently long time, as discussed below in more details.

7.3 Energy management through high-pass filtering

Among the existing strategies to manage the energy content of the reference signal we focus on a high-pass filtering method described in [23], that we denote as the *Borsche filter*. The idea is to subtract the running average the frequency deviation computed over a period a from the instantaneous frequency deviation.

$$\Delta f^e(t) = \Delta f(t) - \frac{1}{a}\int_0^a \Delta f(t-s)\,ds \tag{7.6}$$

Let's note that this Borsche filter takes a very similar form to the Stop/Rate aggregate model. Therefore, the filtering process will impact the network similarly as an homogeneous group of ECLs controlled with the Stop or Rate policy and whose loads have a natural run time $T_n = a$. Therefore, this advanced power control makes sense only for loads which have a shorter run times than the selected parameter a.

There are several advantages at using such a filter. Firstly, the computation of $\Delta f^e(t)$ may easily be decentralized. Secondly, the energy bounds can be theoretically computed. Finally, it allows for error rejection. Indeed, possible measurement bias and error will not propagate in the integral of $\Delta f^e(t)$.

The energy content of the filtered frequency deviation $\Delta f^e(t)$ is bounded by the integral of the non-filtered part in the time window $[t-a, t]$. Let us define $\Delta F(t)$ as the primitive of $\Delta f(t)$. Let us assume $\Delta F(t < 0) = \Delta f(t <$

$0) = 0$.

$$E_f(t,a) = \int_0^t \Delta f^e(s)ds = \Delta F(t) - \frac{1}{a}\int_0^a \Delta F(t-s)ds \qquad (7.7)$$

As $\Delta F(t-s) = \Delta F(t) - \int_0^s \Delta f(t-z)dz$, this gives

$$E_f(t,a) = \int_{t-a}^t \Delta f^e(s)ds = \frac{1}{a}\int_0^a \left(\int_{t-s}^t \Delta f(z)dz\right) ds \qquad (7.8)$$

At time t, the integral of the filtered signal is completely defined by the frequency deviations value in the time window $[t-a,t]$. The energy content $E_f(t,a)$ is normalized by deviation $\phi = 200mHz$ and the result is bounded below by $c_-(a)$(negative frequency deviations) and above by $c_+(a)$(positive frequency deviations).

$$c_+(a) = \max_t \frac{E_f(t,a)}{\phi} \qquad (7.9)$$

$$c_-(a) = \max_t -\frac{E_f(t,a)}{\phi} \qquad (7.10)$$

The energy bound of the filtered frequency signal $c(a) = \max\left(c_-(a), c_+(a)\right)$ can be computed from historical data. Figure 7.2 represents the bounds of the energy content $c_-(a)$ and $c_+(a)$ of the filtered signal and compares it to the maximum energy content of the original frequency signal computed on the same period a. These are computed from 10s-based frequency measurements covering a one year period between Nov. 2014 and Oct. 2015 (Source: RTE).

Bounds of the (non-filtered) frequency integral

The filtered frequency guarantees the energy bounds $c(a)$. However, the original frequency has an energy content whose bounds do not appear on the chart. A physical limit exists thanks to the *time control process* (chapter 10 of [52]). The UCTE operation handbook [145] was more precise: the integral of the frequency deviations should be at most of ± 3000 cycles. The objective of this bound is to insure that frequency-synchronized clocks would experience a maximum shift of 1 minute. Indeed, 50 cycles are counted by clocks as corresponding to a 1 second as the frequency is 50Hz. No compensating action should be undertaken by the system operator while the time shift stays below ± 20 seconds (± 1000 cycles). In normal situations, the target is to keep a shift ± 30 seconds (± 1500 cycles). The normalized deviations $r(t) = \Delta f(t)/\phi$ are therefore subject to the following bounds.

$$\left|\int \Delta f(t)dt\right| \leq 3000 \qquad \Leftrightarrow \qquad \left|\int r(t)dt\right| \leq 15000s = 4h10. \qquad (7.11)$$

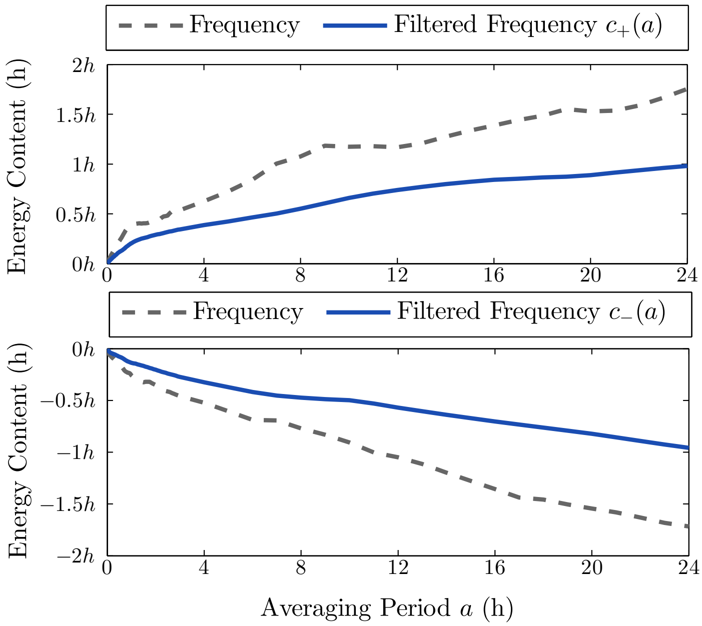


Figure 7.2: Extreme energy content of the frequency deviations (dashed) and bounds of the filtered counterpart (plain) as function of averaging period a in the symmetric case, as observed from 1-year consecutive data. Top: maximum. Bottom: minimum.

As can be seen, thanks to the filter, the theoretical bound c(a) stays much lower than the extreme case (4h10), and even from the normal operation case (2h05). It will therefore limit the energy storage requirement on the group-side.

Asymmetric service

For the sake of generality, it is also important to look at these energy contents in case loads are restricted to deliver an asymmetric flexibility. The consecutive limits are respectively $c_-^{as}(a)$ and $c_+^{as}(a)$. Practically, they do not differ from the symmetric case. Extreme frequency events at the source of the largest energy deviations are by nature asymmetric. It is still interesting to see that this general asymmetric trend can last for very long periods (up to a day). Indeed, even for a reaching 24h, the asymmetric (dash-dot) and symmetric frequency integral presented on figure 7.3 are close to each other.

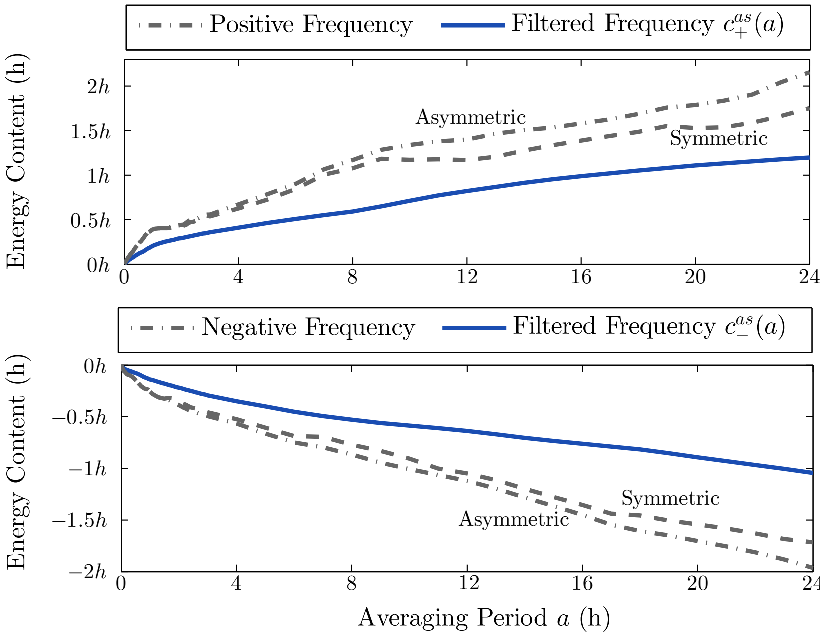


Figure 7.3: Extreme energy content of the frequency deviations (dashed) and bounds of the filtered counterpart (plain) as function of averaging period a in the asymmetric case, as observed from 1-year consecutive data. Top: maximum. Bottom: minimum.

Selection of the averaging period

Let us insist on the terminology. The *response error* or *control error* is the error of the group demand w.r.t. its set-point. We denote by *recovery error* the response error consecutive to the use of an energy management process. On the other had, the *rebound error* is the natural response error that occurs when the energy rebound are not managed at all (cfr. previous chapters).

The optimal choice of a must integrate technical as well as economic considerations. On the economic side, increasing a inevitably leads to larger energy bounds $c(a)$. This supposes that larger portion of the available capacity will be dedicated to rebound management (to avoid the rebound error). This portion of the capacity cannot be sold to the system operator. On the technical side, the averaging period a should be chosen as large as possible to decrease the recovery error. This error should approach the 5%

tolerance band, as requested by ENTSO-E.

As shown on figure 7.4, the best choice seems to be $a = 3h$. Indeed, the *relative* energy bound $c(a)$ starts being more or less constant from $a = 3h$ and larger. The *derivative* is indicated approximately by the red lines on figure 7.4 to highlight this trend. This means that loads with a natural run time T_n equal or higher than $3h$ controlled under the simple Stop/Rate scheme (no filtering process, etc.) would exhibit a *rebound* error of the same power magnitude. It would therefore be counterproductive to exploits loads with $T_n > 3h$ in an advanced control scheme whose filtering process is fixed to $a \leq 3h$.

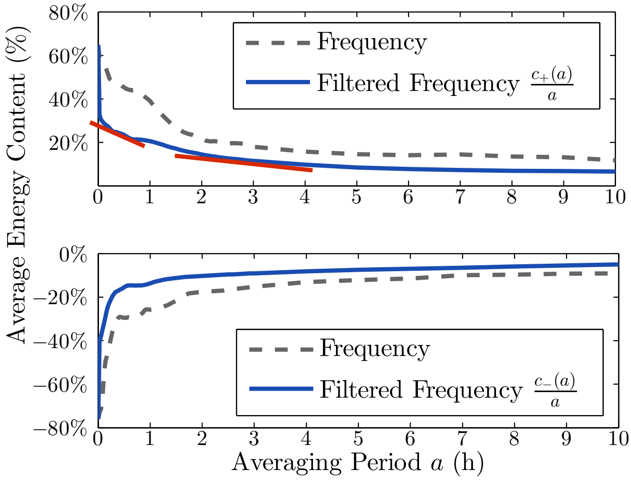


Figure 7.4: Relative energy bounds as a function of parameter a.

Stricter energy bound from pre-qualification tests

The above derived bounds $c(a = 3h)$=20min35 introduced by the filtering process will be used below as basis for the capacity allocation process. This is valid only if the system operator does not require loads to be able to shift more energy *without exhibiting rebound error* than the energy content of the filtered historical frequency deviation.

In many systems, FCR participants must go through pre-qualification tests that are usually more strict, as is the case e.g., in German systems. The energy content of the prequalification profiles in the German case is equal to $30min$[1]. This profile is represented on figure 7.5 for illustration. In such case, the energy bound c exploited in the capacity allocation process would therefore have to be set *artificially* to the energy content of pre-qualification profile, as it is above $c(a = 3h)$, and that load needs to follow this profile without exhibiting rebound error.

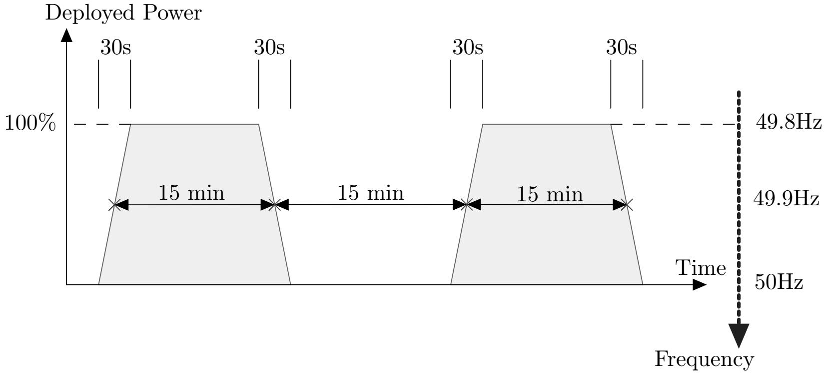


Figure 7.5: Prequalification test in the German System.

[1]See the profile on the document *Model protocol as evidence of the primary control reserve activation.* `https://www.regelleistung.net/ext/static/prequalification`

7.4 Capacity allocation: Perfect Tracking Feasibility set

This section is dedicated to the exact tracking of bounded reference signal.

$$-K_D \leq K_D r^e(t) \leq K_D \qquad - K_D c(a) \leq \int K_D r^e(t) dt \leq K_D c(a)$$

For convenience, the above system is normalized and transformed into a single input $p(t) = P(t)/P_n$, single output $y(t) = D(t)/(\lambda P_n)$ and single state $\tau(t) = T^{dyn}(t)$ system. Note that the input is dimensionless, the output and states have the same dimension (i.e. time). It takes the following form.

$$\mathcal{S} = \begin{cases} \dfrac{d\tau(t)}{dt} = 1 - \dfrac{p(t)}{p_t(\tau(t))} & (7.12) \\ y(t) = \tau(t)p(t) & (7.13) \end{cases}$$

with : $p(t \leq 0) = \nu(t) \in [p_L, p_H]; \tau(0) = \{\tau_0 | \int_{-\tau_0}^{0} \nu(s) ds = T_n\}; \;\; 0 < p_L < 1 < p_H < \infty$.

$$\int_{t-\tau(t)}^{t} p(s)\, ds = T_n \tag{7.14}$$

Without control intervention, the natural input is constant $p = 1$. The energy constraint recalled on equation 7.14 shows that the natural state value is $\tau = T_n$ and therefore, natural output is $y = T_n$.

7.4.1 Consequences of input bounds.

The objective of this section is to prove that there exists an input signal $p(t) \in [p_L, p_H]$ $\forall t > 0$ that verifies the following, for some given value of $0 < K_D < 1$.

$$y(t) = K_D(c(a), T_n, p_L, p_H) r^e(t, a)$$

We want to define the feasible K_D given the run time T_n, power limits p_L and p_H, the filter parameter a and related bounds $c(a)$ of $r^e(t)$. This goes together with showing that the system is stable and is able to store a limited amount of energy $E_s(t)$. The four propositions below show the feasibility and limitations of the control problem.

- Proposition 1 shows that, given any past input $\nu(t)$ belonging to the bounded interval $[p_L, p_H]$, imposing a constant input from time $t = 0$ leads the system to stabilize to an equilibrium point in finite time.

- Proposition 2 defines the concept of *remaining energy* $E_r(t)$. It is an equivalent state of the system that may only evolve when the output $y(t)$ differs from its natural value T_n. It is therefore directly linked with the stored energy $E_s(t)$.

- Proposition 3 states that, given some past bounded inputs, it is always possible to set the output at its natural value $y(t) = T_n$ while ensuring future inputs to stay withing their previous bounds. Furthermore, the input $p(t)$ converges asymptotically to a constant value which is inversely proportional to the *remaining energy* of the group at the moment of the switch (i.e., kept unchanged afterwards as $y(t) = T_n$, proposition 2).

- Proposition 4 argues that input bounds imply bounds on the remaining energy and therefore on the energy storage potential of the group $E_s(t) = -\int (y(t) - T_n)\,dt$.

Proposition 1. *Let* (p_e, τ_e, y_e) *be an equilibrium point of the system* $(p(t), \tau(t), y(t))$*. For any (time) constant* $T_n \in \mathbb{R}_+ \setminus \{0\}$*, any previous bounded inputs* $p(t) = \nu(t), \forall t \in [-\tau_0, 0)$ *s.t.* $\nu(t) \in [p_L, p_H]$ *and the associated initial time delay* τ_0 *s.t.*

$$\{\tau_0 | \int_{-\tau_0}^{0} \nu(s)\,ds = T_n\}$$

the system always reaches the equilibrium $(p_e, \tau_e = T_n/p_e, y_e = T_n)$ *if the input is set to the constant value* $p(t) = p_e(\forall t \geq 0)$*. The equilibrium is reached in finite time* $t_e \leq T_n/p_e$*.*

Proof. Firstly, let us see that the time derivative of the delay $\tau(t)$ is bounded. From equation (7.12) and considering the bounds on the input $p(t)$, we have,

$$(1 - \frac{p_H}{p_L}) \leq \dot{\tau}(t) \leq (1 - \frac{p_L}{p_H}) < 1 \quad \forall t \tag{7.15}$$

Secondly, the time constant T_n and the bounded input function $p(t)$ imply bounds on $\tau(t)$. Indeed, equation (7.14) has to be verified for all time t and all values of $p(t)$. Extreme values of the integral of $p(t)$ define extreme values of $\tau(t)$.

$$\int_{t-\tau_L}^{t} p_L d\tau = T_n \quad \& \quad \int_{t-\tau_H}^{t} p_H d\tau = T_n \quad \forall t \tag{7.16}$$

This gives therefore limits on $\tau(t) \in [\tau_L, \tau_H]\,\forall t$ s.t,

$$0 < \frac{T_n}{p_H} = \tau_L < T_n < \tau_H = \frac{T_n}{p_L} < \infty \tag{7.17}$$

Consequently, as $y(t) = p_e\tau(t)$, the output signal is bounded for all time $t \geq 0$.

$$0 < \frac{T_n p_L}{p_H} \leq y(t) \leq \frac{T_n p_H}{p_L} < \infty \quad \forall t \tag{7.18}$$

The system is therefore stable in the *Bounded-input Bounded-output* sense for any input $p(t) \in [p_H, p_L]$.

Starting from $t = 0$, we have $\tau(t) > t$ ($\forall t < t_e$) before a threshold time t_e. As $\dot{\tau}(t) < 1$ we can derive that $\exists t_e$ such that $\tau(t_e) = t_e$.

Equation (7.14) expressed at threshold time t_e gives the following.

$$\int_{t_e-\tau(t_e)}^{t_e} p(s)ds = \int_0^{t_e} p(s)ds = p_e t_e = T_n \tag{7.19}$$

Therefore, the threshold time is $t_e = T_n/p_e < \infty$ and defines the equilibrium delay/state $\tau(t_e) = t_e = \tau_e$. Indeed, the time derivative of $\tau(t)$ is zero for all time after $t = t_e$ (eq. (7.12)). Indeed, $p(t) = p(t - t_e) = p_e$ ($\forall t \geq t_e$). The equilibrium output may also be computed by $y_e = \tau_e p_e = T_n$. This result is crucial and shows that steady-state may only be reached when the output is constant and equal to its natural value T_n. □

Proposition 2. *Considering system $\mathcal{S}$, the remaining energy function $E_r(t)$ is a positive definite function.*

$$E_r(t) = \int_{t-\tau(t)}^{t} \left(T_n - \int_\theta^t p(s)\,ds\right) d\theta \quad > 0 \tag{7.20}$$

$$= T_n\tau(t) - \int_{t-\tau(t)}^{t} \left(\int_\theta^t p(s)ds\right) d\theta \tag{7.21}$$

$$= \int_{t-\tau(t)}^{t} \left(\int_{t-\tau(t)}^{\theta} p(s)ds\right) d\theta \tag{7.22}$$

The remaining energy $E_r(t)$ is conserved if and only if $y(t) = T_n$. It consists in the amount of energy that remains to be consumed by all running loads before those will stop.

Proof. The energy function $E_r(t)$ is trivialy positive definite from equation (7.22) as $p(t) > 0, \forall t$, and $\tau(t) > 0, \forall t$. Let us denote by $I_p(t)$ is the primitive of $p(t)$ at time t and define

$$b(t,\theta) = \int_\theta^t p(s)\,ds = I_p(t) - I_p(\theta)$$

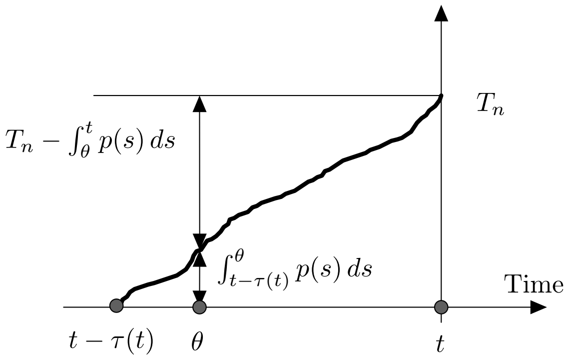


Figure 7.6: The remaining energy of the group $E_r(t)$ corresponds to the integral under the curve shown in the graph and situated in the interval $[t-\tau(t), t]$. This curve at a certain time θ is the remaining time $T_r(\theta)$ of a load that has arrived at that time and was subject to power control $p(s), \forall s \in [\theta, t]$. If the load continues running from time t at its natural rate (i.e., $p = 1$), it would stop at time $t + T_r(\theta)$. The curve represents this time $T_r(\theta) = T_n - \int_\theta^t p(s)\,ds$. The reference axis can be understood as *sliding rightward* together with time t. This is why the origin is t. The curve is different at each time t but always encounters the y-axis at T_n and slides down from this point backward in time.

We compute below the time derivative $\dot{E}_r(t)$ of the remaining energy.

$$\begin{aligned} \dot{E}_r(t) &= T_n\dot{\tau}(t) - \frac{d}{dt}\left(\int_{t-\tau(t)}^{t} (\int_\theta^t p(s)\,ds)\,d\theta\right) && (7.23)\\ &= T_n\dot{\tau}(t) - \frac{d}{dt}g(t) && (7.24)\end{aligned}$$

Let's develop $g(t)$.

$$\begin{aligned} g(t) = \int_{t-\tau(t)}^{t} b(t,\theta)d\theta &= I_p(t)\int_{t-\tau(t)}^{t} d\theta - \int_{t-\tau(t)}^{t} I_p(\theta)\,d\theta && (7.25)\\ &= \tau(t)I_p(t) - \int_{t-\tau(t)}^{t} I_p(\theta)\,d\theta && (7.26)\end{aligned}$$

Therefore, recalling that $I_p(t) - I_p(t-\tau(t)) = \int_{t-\tau(t)}^{t} p(s)ds = T_n$ and $\dot{I}_p(t) = p(t)$,

$$\begin{aligned} \frac{d}{dt}g(t) &= \dot{\tau}(t)I_p(t) + \tau(t)\dot{I}_p(t) - I_p(t) + I_p(t-\tau(t))(1-\dot{\tau}(t)) && (7.27)\\ &= T_n(\dot{\tau}(t) - 1) + \tau(t)p(t) && (7.28)\\ &= T_n(\dot{\tau}(t) - 1) + y(t) && (7.29)\end{aligned}$$

Therefore,

$$\dot{E}_r(t) = T_n - y(t) \tag{7.30}$$

We have that $\dot{E}_r(t) = 0$ whenever $y(t) = T_n$. In the uncontrolled setting ($p(t) = 1, \tau(t) = T_n, \forall t$), the remaining energy takes its natural value $E_n^r = T_n^2/2$. The remaining energy is linked to the energy stored by the group $E_s(t) = -\int (y(t) - T_n)\, dt$,

$$E_r(t) = \frac{T_n^2}{2} + E_s(t) = E_n^r + E_s(t)$$

□

Proposition 3. *Considering the system $\mathcal{S}$: for any bounded past input $p(t < 0) = \nu(t) \in [p_H, p_L]$ and related outputs, it is always possible to maintain future output to its natural level $y(t) = T_n$ $\forall t \geq 0$ while ensuring inputs to stay within their initial bounds (i.e. bounds of $\nu(t)$). Furthermore, the system $\mathcal{S}$ converges asymptotically toward an equilibrium ($p_e, \tau_e = T_n/p_e, y_e = T_n$).*

Proof. Assuming $y(t) = p(t)\tau(t) = T_n$ $\forall t > 0$ and replacing T_n from equation (7.14) gives,

$$p(t) = \frac{1}{\tau(t)} \int_{t-\tau(t)}^{t} p(s)ds \quad \forall t > 0 \tag{7.31}$$

The input $p(t > 0)$ corresponds to the average of its (bounded) previous values computed on the rolling *time window* $[t - \tau(t), t)$ of strictly positive width $\tau(t)$. Therefore,

$$p_H \leq p(t) \leq p_L \quad \forall t \tag{7.32}$$

Future inputs are bounded if the output is set to its natural value. Furthermore, the input $p(t)$ and time delay $\tau(t)$ are asymptotically converging to constant values. Indeed, the system is asymptotically stable in the sense of Lyapunov. Let us first define the positive semi-definite function $V(\tau, p, t) \geq 0$.

$$\begin{aligned}
V(\tau, p, t) &= \int_{-\tau(t)}^{0} \big(p(t) - p(t+s)\big)^2 ds && (7.33)\\
&= p^2(t)\tau(t) + \int_{-\tau(t)}^{0} p^2(t+s)ds - 2p(t)\int_{-\tau(t)}^{0} p(s)ds && (7.34)\\
&= \int_{-\tau(t)}^{0} p^2(t+s)ds - T_n p(t) && (7.35)
\end{aligned}$$

The function V is zero only if the input stays constant in the time window $[t-\tau(t), t]$. Furthermore, its time derivative $\dot{V}(\tau, p, t)$ is negative semi-definite.

$$\dot{V}(\tau, p, t) = \left(\frac{d}{dt} \int_{t-\tau(t)}^{t} p^2(s)\, ds \right) - T_n \dot{p}(t) \tag{7.36}$$

As $y(t) = \tau(t)\, p(t)$ is set to T_n, we have $\tau(t) = T_n/p(t)$ and $\dot{\tau}(t) = -T_n \dot{p}(t)/p^2(t)$ such that $T_n \dot{p}(t) = -p^2(t)\dot{\tau}(t)$. Therefore,

$$\dot{V}(\tau, p, t) = \Big(p^2(t) - p^2(t-\tau(t))(1-\dot{\tau}(t)) \Big) + p^2(t)\dot{\tau}(t) \tag{7.37}$$

$$= p^2(t) \left[(1 - \frac{p^2(t-\tau(t))}{p^2(t)}) + \dot{\tau}(t)\, \big(1 + \frac{p^2(t-\tau(t))}{p^2(t)}\big) \right] \tag{7.38}$$

Recalling that $\dot{\tau}(t) = 1 - \dfrac{p(t)}{p(t-\tau(t))}$, we have that $\dfrac{p^2(t)}{p^2(t-\tau(t))} = \big(1 - \dot{\tau}(t)\big)^2$.

$$\dot{V}(\tau, p, t) \quad = \quad \frac{p^2(t)}{(1-\dot{\tau}(t))^2} \left[(1-\dot{\tau}(t))^2 - 1 + \dot{\tau}(t)(1-\dot{\tau}(t))^2 + \dot{\tau}(t) \right] \tag{7.39}$$

$$= \quad \left(\frac{p(t)\dot{\tau}(t)}{(1-\dot{\tau}(t))} \right)^2 (\dot{\tau}(t) - 1) \quad \leq \quad 0 \tag{7.40}$$

Indeed, $(\dot{\tau}(t) - 1)$ is always negative, as equation (7.1) shows, and the condition $p(t) > 0$ insures $\dot{\tau} < 1$. Consequently, the function $V(\tau, p, t)$ must decrease and tend asymptotically to zero as soon as $y(t) = T_n$. So is it for the time derivative $\dot{V}(\tau, p, t)$. This implies that $\dot{\tau}(t)$ will also tend to zero. As $\dot{\tau}(t) = 0$ implies $p(t) = p(t-\tau(t))$, steady-state is reached if and only if $p(t) = p(t-\tau(t))$ for all successive time t, meaning that the input will be constant. □

Proposition 4. *The asymptotic equilibrium reached in proposition 3 can be computed from the value of the remaining energy E_0 at time $t = 0$.*

$$(p_e = T_n^2/2E_0, \tau_e = 2E_0/T_n, y_e = T_n)$$

The input bounds imply bounds on E_0. Consequently, the energy that the group is able to store is also bounded.

$$-\frac{T_n^2}{2} < \frac{T_n^2}{2}\frac{1-p_H}{p_H} \leq E_s(t) \leq \frac{T_n^2}{2}\frac{1-p_L}{p_L}$$

The remaining energy is positive definite, which explain the lower bound $-T_n^2/2 = -E_n^r$.

Proof. From proposition 3, $p(t)$ converges to a constant value p_e for any past bounded input $\nu(t)$ if $y(t)$ is set to T_n while the system converges to the equilibrium point $(p(t) = p_e, \tau(t) = T_n/p_e, y(t) = T_n)$. Proposition 2 showed that the stored and remaining energy are conserved for $t > 0$ as $y(t) = T_n$. The remaining energy E_0 at time $t = 0$ is used to compute the asymptotic equilibrium, with $\int_{-\tau_0}^{0} \nu(s)ds = T_n$.

$$E_0 = T_n\tau_0 - \int_{-\tau_0}^{0} (\int_{\theta}^{0} \nu(s)ds)d\theta \tag{7.41}$$

At the asymptotic equilibrium, the remaining energy is defined by equation (7.22) in which we replace $\tau(t) = \tau_e$ and $p(t) = p_e = T_n/\tau_e$. This remaining energy is equal to E_0 as $y(t)$ was kept to T_n.

$$T_n\tau_e - p_e\frac{\tau_e^2}{2} = E_0 \tag{7.42}$$

$$\frac{T_n\tau_e}{2} = E_0 \quad \Leftrightarrow \quad \tau_e = \frac{2E_0}{T_n} \quad \Leftrightarrow \quad p_e = \frac{T_n^2}{2E_0} \tag{7.43}$$

Bounds on inputs (i.e., applying to p_e) imply bounds on E_0, and therefore on $E_r(t)$ (or $E_s(t)$) as the only constraint on the initial input $\nu(t)$ is that it must be bounded.

$$\frac{T_n^2}{2p_H} \leq E_r(t) \leq \frac{T_n^2}{2p_L} \quad \Leftrightarrow \quad \frac{T_n^2}{2}\frac{1-p_H}{p_H} \leq E_s(t) \leq \frac{T_n^2}{2}\frac{1-p_L}{p_L} \tag{7.44}$$

Furthermore, physical limits on the upper bound $p_H < \infty$ (or positive definite nature of of $E_r(t)$) imply $-T_n^2/2 < E_s(t)$. Intuitively, this means that loads cannot anticipate more energy than what was available at the origin, when the group was in its natural equilibrium. □

Remark 1. *The power limits give a first couple of constraint on the symmetric droop K_D. These limits are relative to upward and downward regulation. They are respectively denoted $K_D^{p,up}$ and $K_D^{p,dn}$.*

$$K_D \leq K_D^{p,up} = 1 - p_L \qquad \& \qquad K_D \leq K_D^{p,dn} = p_H - 1 \tag{7.45}$$

Remark 2. *The reference signal $r^e(t)$ followed by the group's demand change $x(t) = y(t)/T_n - 1$ must have a bounded energy content which corresponds to the maximum amount of the energy that can be stored by the group. We must impose the following:*

$$\max_t |E_s(t)| = K_D c(a) \tag{7.46}$$

Accordingly, the remaining energy E_0 is bounded and so is the power p_e at the asymptotic equilibrium.

$$T_n - 2K_D c(a) \leq \frac{2E_0}{T_n} \leq T_n + 2K_D c(a) \tag{7.47}$$

$$\Rightarrow \frac{T_n}{T_n + 2K_D c(a)} \leq p_e \leq \frac{T_n}{T_n - 2K_D c(a)} \tag{7.48}$$

This gives a second couple of limits on the symmetric droop that can be offered by the group.

$$K_D \leq K_D^{e,up} = \frac{T_n}{2c(a)} \frac{1-p_L}{p_L} \quad \& \quad K_D \leq K_D^{e,dn} = \frac{T_n}{2c(a)} \frac{p_H - 1}{p_H} \tag{7.49}$$

We now have a first energy related bound on K_D that depends on the energy management filter parameter a and the related energy $c(a)$ of the filtered frequency signal $\Delta f^e(t)$.

Remark 3. *Before reaching its equilibrium point, the input $p(t)$ had to fluctuate around it. This require to define stricter bounds than K_D^e. The above bounds K_D^e do not take dynamical considerations into account. The group could be asked to provide its flexibility $1 \pm K_D$ for an additional and infinitesimal amount of time. This worst case scenario is illustrated on figure 7.7 and results in the following relations.*

$$p_L \leq \frac{T_n(1-K_D)}{T_n + 2K_D c(a)} \quad \& \quad \frac{T_n(1+K_D)}{T_n - 2K_D c(a)} \leq p_H \tag{7.50}$$

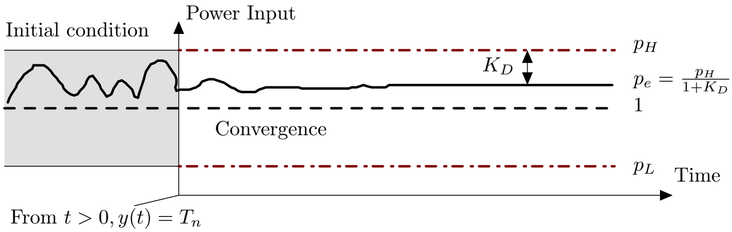


Figure 7.7: Worst Case scenario (downward regulation, demand increase) defining the bound $K_D^{dp,dn}$ on the gain K_D. The initial condition has resulted in shifting an energy $K_D c(a)$. From time t>0, the output is set to $y(t) = T_n$, and the group had reaches the corresponding steady-state $p_e(E_0 = K_D c(a))$ in the convergence period. The worst case is simple: the group must still be able to impose an very brief change $y = T_n(1 + K_D)$ to its output while respecting the input bound p_H.

This gives the following limits.

$$K_D \leq K_D^{dp,up} = \frac{(1-p_L)T_n}{2c(a)p_L + T_n} \quad \& \quad K_D \leq K_D^{dp,dn} = \frac{(p_H - 1)T_n}{2c(a)p_H + T_n} \tag{7.51}$$

Figure 7.8 shows the acceptable values for K_D given by all limits derived in (7.49) and (7.51). The individual load power limits are symmetric such that $1 - p_L = p_H - 1 = \Delta p$. On the right chart, $\Delta p = 50\%$. On the left chart $\Delta p = 80\%$. The bounds KD^{dp} are the most constraining limit. The chosen frequency filter is such that $a = 3h$ and $c(3h)$ =20min35.

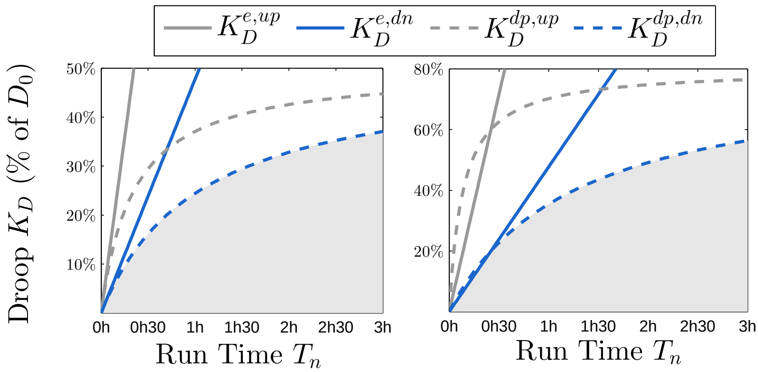


Figure 7.8: Acceptable Droop K_D in function of the load run time and according to all derived limits. Left chart: $1 - p_L = p_H - 1 = 50\%$. Right chart: $1 - p_L = p_H - 1 = 80\%$.

7.4.2 Empirical evidence: input bounds are respected

The power profile that needs to be followed by the group, namely $K_D r^e(t)$, has potentially dynamical effects that are not captured in the above bounds. We would like to assess if the above derived bounds are coping for the potential dynamical effects resulting in higher than expected inputs, consecutive to output signals with identical energy content but of different shapes. We conduct therefore a large number of tests with different profiles χ of *the same amplitude* K_D but different *shapes* and energy *content*. These are illustrated on figure 7.9: sinusoidal, square and impulses, alternate square, ramp and polynomials.

We impose each profile as reference in an inverse model. This process allows to see the extreme input values that were required to perfectly follow the tested reference χ.

Each profile χ_i is dimensionless. The output must therefore follow $y(t) = T_n\chi_i$. The integral of χ_i has the dimension of time and is bounded by a known value $K_D T_i$. In the simulations, we consider all above profiles and some other related ones. For instance we test all possible square cycles

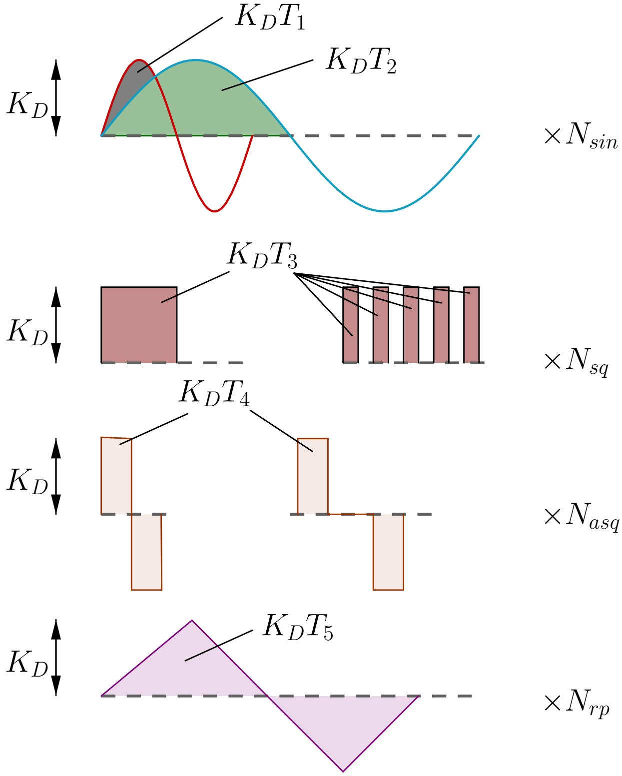


Figure 7.9: Illustration of the tested profiles (dimensionless).

(i.e. we vary the time at $\pm K_D$ as well as the time spent at zero). We also consider profiles independently or repeated several times in a row (e.g., T_i is adapted necessary).

For a certain profile χ_i we will therefore observe the experimental power limits $p_{L,i}^{exp}$ and $p_{H,i}^{exp}$ that are reached in the simulations. They are supposed to stay within the theoretical bounds $p_L^{th}(T_i/T_n, K_D)$ and $p_H^{th}(T_i/T_n, K_D)$, that are both functions of the relative energy content T_i/T_n and the maximum amplitude K_D of the reference profile.

$$p_L^{th}(T_i/T_n, K_D) = \frac{1-K_D}{1+2K_D\dfrac{T_i}{T_n}} \qquad \& \qquad p_H^{th}(T_i/T_n, K_D) = \frac{1+K_D}{1-2K_D\dfrac{T_i}{T_n}}$$

We need to insure the following $\forall i, K_D, T_n$

$$p_L^{th}(T_i/T_n, K_D) \le p_{L,i}^{exp} < p_{H,i}^{exp} \le p_L^{th}(T_i/T_n, K_D)$$

Let's define the *relative margin* to bound e_L and e_H, respectively for upward and downward regulation. These margins are the relative difference between the most extreme observations and the theoretical bound of the power input. They are both defined such that they should ideally be positive (i.e., the observed extreme input never exceeded the theoretical bounds) and be close to zero (i.e., the theoretical bound is tight).

$$e_L = \frac{p_{L,i}^{exp}}{p_{L,i}^{th}} - 1 \tag{7.52}$$

$$e_H = \frac{p_{H,i}^{th}}{p_{H,i}^{exp}} - 1 \tag{7.53}$$

In practice, the group is considered at rest ($y(t) = T_n$) from $t = -\infty$ up to $t = 0$, when the test starts. We test the profiles of fixed parameter T_i (i.e. different K_D are tested) on groups with different run time T_n. Such procedure is fully symmetric to the one consisting in testing signals of different T_i on a unique group.

We can easily prove this symmetry. Let's pose $a(t) = \tau(t)/T_n$ and consider two groups of loads (index 1 and 2) of which demand is imposed to follow two different reference signals $T_{n,1}\chi(t)$ and $T_{n,2}\chi(Zt)$; respectively. Note that the second signal is the time distorted version ($t \to Zt$) of the first one with adapted magnitude. We have to solve the following.

$$S_1 = \begin{cases} a_1(t)p_1(t) = \chi(t) \\ \displaystyle\int_{-T_{n,1}a_1(t)}^{0} p_1(t+s)\,ds = T_{n,1}\ \forall t \end{cases} \quad S_2 = \begin{cases} a_2(t)p_2(t) = \chi(Zt) \\ \displaystyle\int_{-T_{n,2}a_2(t)}^{0} p_2(t+s)\,ds = T_{n,2}\ \forall t \end{cases}$$

Replacing the first equation in the second gives the following.

$$\int_{-T_{n,1}\chi(t)/p_1(t)}^{0} p_1(t+s)\,ds = T_{n,1}\ \forall t \qquad \int_{-T_{n,2}\chi(Zt)/p_2(t)}^{0} p_2(t+s)\,ds = T_{n,2}\ \forall t$$

We now want to compute the integral of $p_1(Zt+Zs)$ defined in the same bounds than the left equation. As both equations are valid at all time t, we can replace t by Zt in the integrand *and the boundary value.* Furthermore, posing $v = Zs$ gives $ds = dv/Z$ and adapted bounds leading to the following.

$$\int_{-ZT_{n,1}\chi(Zt)/p_1(Zt)}^{0} p_1(Zt+v)\,dv = ZT_{n,1} \qquad \int_{-T_{n,2}\chi(Zt)/p_2(t)}^{0} p_2(t+s)\,ds = T_{n,2}$$

We can therefore conclude below.

$$ZT_{n,1} = T_{n,2} \qquad \Leftrightarrow \qquad p_1(Zt) = p_2(t) \tag{7.54}$$

Remark 4. *Let's consider a group of loads with run time T_n. The group's demand is required to follow a distorted version $T_n\chi(Zt)$ of a reference $\chi(t)$ which leads the power input to reach some extreme values. We conclude from the above developments that the exact same extreme values would have been observed if loads with run time T_n/Z had to follow the demand reference $T_n\chi(t)/Z$.*

Test Results

The figures 7.10 and 7.11 present the results of all conducted tests.

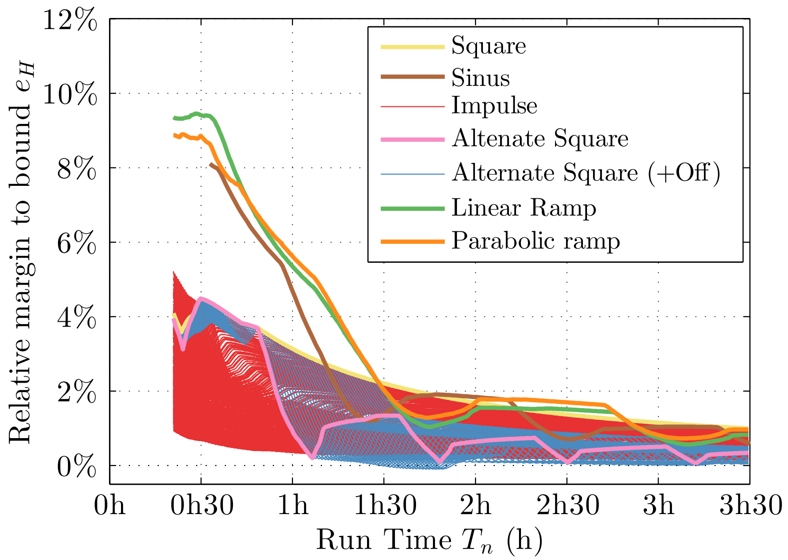


Figure 7.10: Relative bound to margin for downward regulation e_H (demand increase).

We can firstly observe that certain profiles are more likely than other to push the inputs close to their bounds. This suggests that some dynamical effects take place. This is particularly true for alternate squares and impulse signals.

Also, we observe the following elements.

1. Loads with small run time: $K_D T_i >> T_n$. The bounds are respected for all signals. Indeed, the minimum value of both relative margins to bound e_H and e_L are always well above zero.

2. Large run time: $K_D T_i < T_n$. The bounds seems asymptotically perfect as $T_n \to +\infty$.

3. Moderate run time: $K_D T_i \simeq T_n$. The inputs do not always respect their theoretical bound. Yet, they stay in the worst case very close to it ($< 0.2\%$ around it).

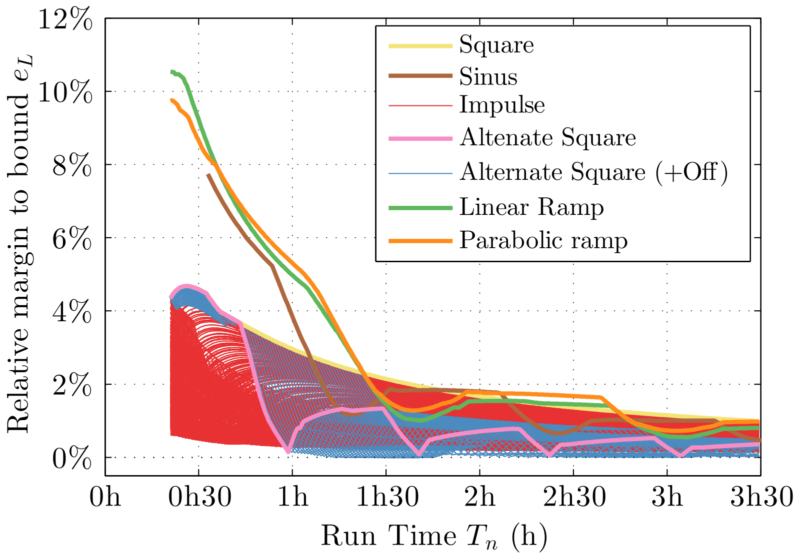


Figure 7.11: Relative bound to margin for upward regulation e_L (demand decrease).

7.4.3 Intuitive explanation of the test results

To understand the test results charts, let's recall that the theoretical bounds are derived from the following concept (stated for demand increase).

Remark 5. *The largest acceptable amplitude K_D of all output signals with bounded energy content $K_D T_i$ is such that, at the moment that a group has been storing an energy $E_s = -K_D T_i$ and had time to reach the corresponding equilibrium $p_e(E_s, T_n)$, loads are are still capable (while respecting their* *power limit) of imposing their power input to go to* $p(t) = p_e(1 + K_D)$ *for an infinitesimal amount of time.*

Let's also recall that the equilibrium p_e is reached asymptotically when the group's output is maintained at equilibrium $y(t) = T_n$ for a sufficiently long time (proven below to be around $3\tau_e$). This process is denoted as the *convergence process.* We have, during the convergence process,

$$p(t) = \frac{1}{\tau(t)} \int_0^{\tau(t)} p(t+s)\, ds$$

This means that, at the moment the convergence process starts, the extreme values of $p(t)$ are situated in its past. These extreme values can either lie within the immediate past $[t - \tau(t), t])$, or have been observed previously in time. In the first case, it would mean that the output reference that had to be followed has been impacting the group mostly in the recent past. There are two possible situations in which this occurs.

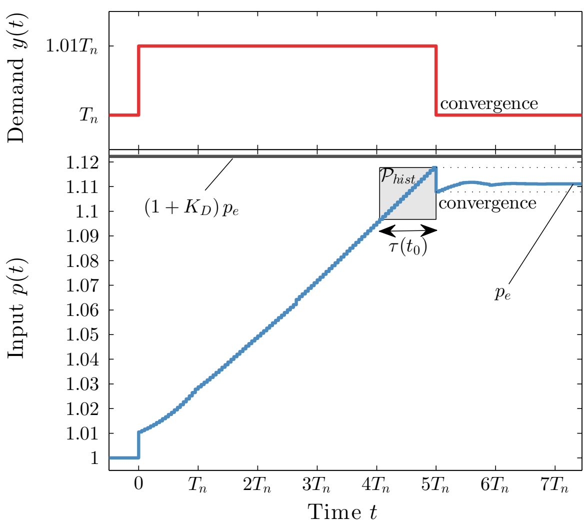


Figure 7.12: Power input $p(t)$ required to follow an output reference signal with large periodicity and modest energy content (demand increase).

The first case occurs for output reference signals whose periodicity is large w.r.t. T_n but whose energy content stays small enough compared to $0.5T_n^2$. This is the case if , e.g., a small and constant offset $(1 + K_D)$ is imposed to the demand of the group for a long duration.

As shown on figure 7.12, the moderate energy content of the output change leads the dynamic run time $\tau(t)$ to stay relatively close to its initial value T_n. During the convergence process, the input evolves as a running average computed on the recent past $[t - \tau(t), t]$. Let's denote by $\mathcal{P}_{hist}$ the set of input values in the recent past $[t_0 - \tau(t_0), t_0]$ *at the start of the convergence period*. The values in $\mathcal{P}_{hist}$ are by far the largest inputs that have been observed. Additionally, their dispersion around their own average is small. More importantly, the second derivative of $p(t)$ is increasing within $\mathcal{P}_{hist}$. Altogether, this means that the asymptotic equilibrium p_e that will eventually be reached is (1) very close to the largest input ever observed and (2) situated closer from the highest values of $\mathcal{P}_{hist}$. Therefore, the bound derived from applying an extra factor $1 + K_D$ to p_e is likely to be slightly above than the actual extreme values of the inputs. The second case is when the output reference has a very small energy content. At the limit, a reference with infinitesimal energy content leads the asymptotic equilibrium to be equal to the natural load's power: $p_e \to 1$. This is illustrated on figure 7.13.

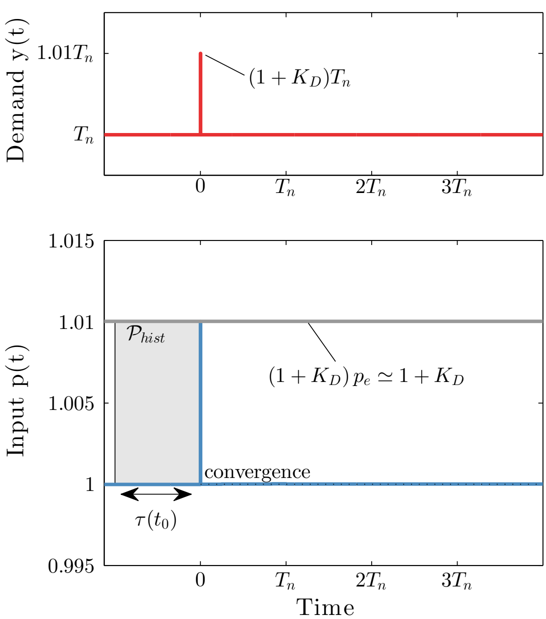


Figure 7.13: Power inputs required to follow an output reference signal with very small energy content (demand increase).

A closer look to alternate square reference profiles

Figure 7.14 presents a detailed view of the results corresponding to alternate square profiles for downward regulation (most constraining).

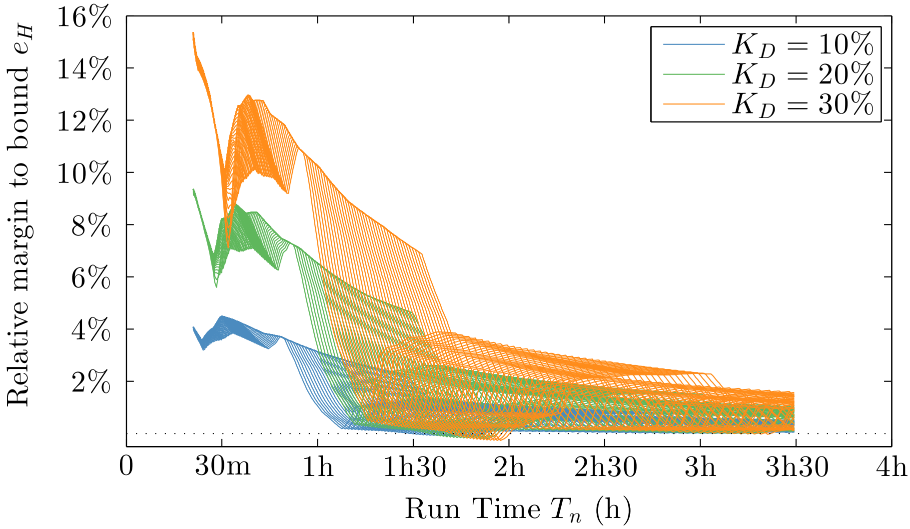


Figure 7.14: Relative margin from observation to theoretical bound with $T_i = c(a = 3h)$ (downward regulation, demand increase).

Alternate square profiles consist in a first positive sequence $\chi(t) = K_D$ lasting for T_i followed by a negative sequence of the same duration. A certain time interval separates both sequences (see on Fig.7.9 the profiles pined with T_4). On figure 7.14, a large number of tests are conducted with varying off-time intervals. These profiles are applied successively with different magnitude K_D to different homogeneous groups of loads of different run time T_n. The maximum energy content $K_D T_i$ of each tested profile χ_i is equal to the energy bound $c(a)$ of the filtered historical frequency of parameter $a = 3h$, that is: $T_i = c(a = 3h)$. This allows to observe which loads could potentially go out of their bounds for this specific filter.

Three elements should be observed. First, the bound seems too strict for loads with small run time. Secondly, the bound is tight for loads with larger run time. Thirdly, the bound seems not strict enough for certain loads of run time in the interval $1h30 - 2h$.

Let's highlight this phenomenon clearly. We impose the group output to follow a reference $\chi(t)$: a succession of 10 alternate squares of height $K_D = 30\%$. The positive and negative sequences are both of known duration T_i and the off-periods of duration $T_i - 1$. On figure 7.15, this profile (top chart) is applied to two different groups in which load run times are respectively $T_n = T_i$ (mid chart) and $T_n = 10T_i$ (bottom chart). The theoretical limit $K_D^{dp,dn}$ (i.e. downward reserve, increase of demand) is shown in both cases.

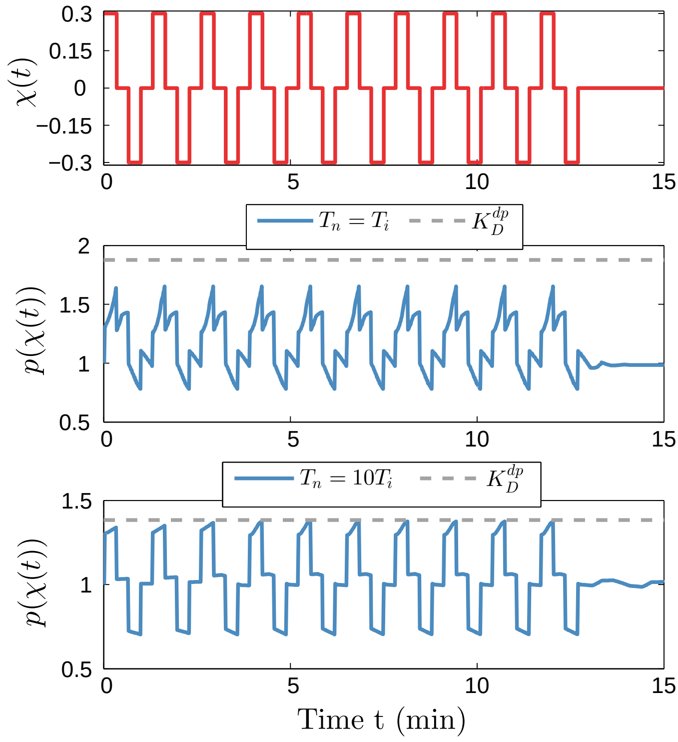


Figure 7.15: Inputs required to follow an alternate square shaped signal of amplitude $K_D = 0.3$ and period $4T_i+2$. Top: reference profile. Middle: power input for $T_n = T_i = 20s$. Bottom: power input for $T_n = 10T_i = 200s$.

Groups for which the maximum energy content of the signal is relatively large compared to their natural remaining energy (mid chart) will never experience inputs getting close to the (energy based) limit $K_D^{dp,dn}$. The inverse is true in case the energy content is small. Indeed, the influence of dynamics and power based consideration are dominating.

7.4.4 Square shape profile and alternative bounds.

Let's impose to a group in steady-state at time $t = 0$ with a zero stored energy $E_s(t)$ to follow a square shaped profile of power magnitude $\pm T_n K_D$ and energy content $\pm T_n K_D c(a)$ after which the group's demand is set to its natural value.

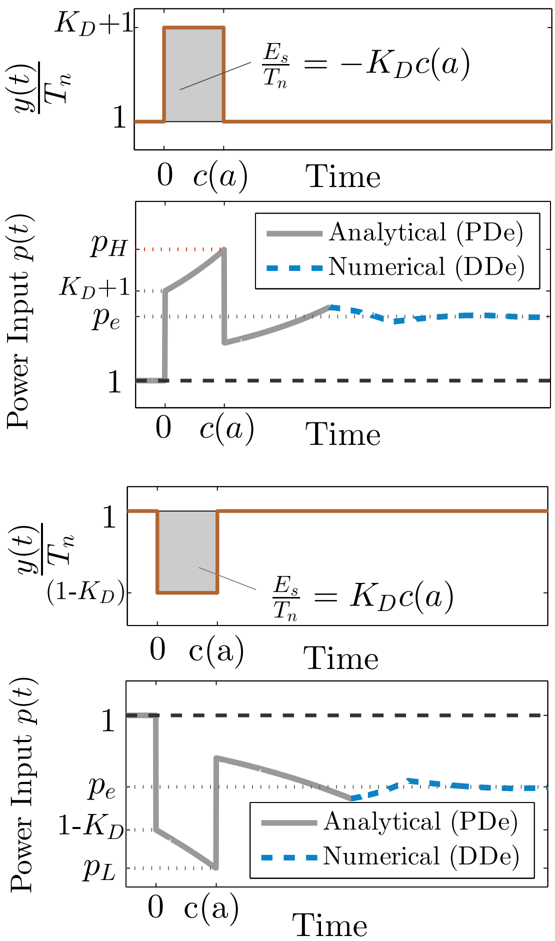


Figure 7.16: Evolution of individual input power $p(t)$ for largest acceptable energy shifting. Left : $1 \leq p(t) \leq p_H$, demand anticipation. Right: $p_L \leq p(t) \leq 1$, demand deferral.

The group must maintain $y(t) = T_n(1 \pm K_D)$ in a time window $t \in [0, \theta]$ with $\theta = c(a)$. The resulting stored energy is $E_s(t > \theta) = \pm T_n K_D c(a)$. The remaining energy of the group is $E_r(t > \theta) = T_n^2/2 \mp T_n K_D c(a)$ (i.e., note the sign inversion). We already know from equation (7.43) that the final equilibrium point will be

$$p_e = \frac{T_n}{T_n \mp 2K_D c(a)} \qquad \& \qquad \tau_e = T_n \mp 2K_D c(a)$$

The system $\mathcal{S}$ (i.e., eq. (7.12) and (7.13)) starts with initial conditions $\nu(t < 0) = 1$ and $\tau_0 = T_n$. In a first phase, the system demand is switched to $(1 \pm K_D)T_n$. This is performed in a certain time window $\forall t \in [0, c(a))$. As long as $\tau(t) < t$, eq. (7.12) can be simplified to eq. (7.55) as $p(t - \tau(t)) = 1$.

$$\dot{\tau}(t) = 1 - p(t) \tag{7.55}$$

$$y(t) = (1 \pm K_D)T_n = p(t)\tau(t) \tag{7.56}$$

The corresponding input $p(t)$ is the solution of a partial differential equation. An analytic solution exists and is shown below. The main branch W_0 of the Lambert function [35] appears in this solution. The Lambert function $w = W_0(x)$ is the solution to the equation: $we^w = x$.

$\forall t \in [0, c(a))$:

$$p(t) = \frac{1}{1 + W_0\left(\frac{\mp K_D}{1 \pm K_D} \exp\left(\frac{\mp K_D}{1 \pm K_D} + \frac{t}{(1 \pm K_D)T_n}\right)\right)} \tag{7.57}$$

After this first period, the power $y(t)$ is switched back to its equilibrium value. Part of the system can be solved analytically if $c(a)$ is such that $c(a) < t < \tau(t)$ (i.e., $p_t(\tau(t)) = 1$). It takes a similar form than (7.57), shown below. Let's denote by $p_s = p(t = c(a)^+) = p(t = c(a)^-)/(1 \pm K_D)$ as the input required to switch the output back to T_n.

$\forall t > c(a), c(a) < \tau(t)$:

$$p(t) = \frac{1}{1 + W_0\left(\frac{1 - p_s}{p_s} \exp\left(\frac{1 - p_s}{p_s} + \frac{t - c(a)}{T_n}\right)\right)} \tag{7.58}$$

We could not derive a closed form solution of the system for $t > \tau(t)$. The rest of the solution is therefore solved numerically.

Maximum square height K_D respecting input limits

As shown on figure 7.16, the extreme values of $p(t)$ are reached just before $y(t)$ is switched back to T_n. We could therefore derive a limit to the amplitude K_D^{sq} that lead the extreme values to be exactly equal to the input bounds.

$$p_L = \frac{1}{1 + W_0\left(\frac{K_{D,p_L}^{sq}}{1-K_{D,p_L}^{sq}}\exp\left(\frac{K_{D,p_L}^{sq}}{1-K_{D,p_L}^{sq}} + \frac{c(a)}{(1-K_{D,p_L}^{sq})T_n}\right)\right)} \tag{7.59}$$

$$p_H = \frac{1}{1 + W_0\left(\frac{-K_{D,p_H}^{sq}}{1+K_{D,p_H}^{sq}}\exp\left(\frac{-K_{D,p_H}^{sq}}{1+K_{D,p_H}^{sq}} + \frac{c(a)}{(1+K_{D,p_H}^{sq})T_n}\right)\right)} \tag{7.60}$$

Equations (7.59) and (7.60) are inverted obtain the bounds K_{D,p_L}^{sq} and K_{D,p_H}^{sq} respective to lower and higher power limits p_L and p_H (the minus sign is relative to p_H).

$$K_{D,p_{L,H}}^{sq} = \pm\frac{W_0\left((1+\frac{c(a)}{T_n})\frac{1-p_{L,H}}{p_{L,H}}\exp\left(\frac{1-p_{L,H}}{p_{L,H}} - \frac{c(a)}{T_n}\right)\right)}{W_0\left((1+\frac{c(a)}{T_n})\frac{1-p_{L,H}}{p_{L,H}}\exp\left(\frac{1-p_{L,H}}{p_{L,H}} - \frac{c(a)}{T_n}\right)\right) + (1+\frac{c(a)}{T_n})} \tag{7.61}$$

7.4.5 A word on the bound strictness

The above derived bounds could be overly restrictive. Let's suppose that the behavior of the 3h run time loads responding with a simple proportional controller (previous chapter) is acceptable on a system perspective. This would mean that the energy rebound their response introduces is not problematic, and that we should allow them to offer their full response capability to the grid (e.g., proportional factor of $K_D^p = min(p_H - 1, 1 - p_L)$) instead of introducing limiting factor from the capacity allocation process (i.e., restricting it to $K_D^{dp} < K_D^p$). Indeed, the artificial limitations have a strong impact on profitability. A simple way to overcome this limit is to scale the above derived bounds such that they become equal to the technical capability (i.e., $1-p_L$ or p_H-1) for run times $T_n \geq 3h$. It is however likely to bring loads with shorter run time out of their power limits.

$$K_D^{dp,scaled}(T_n) = \begin{cases} \frac{K_D^{dp}(T_n)}{K_D^{dp}(3h)}K_D^p & \text{if } T_n < 3h \\ K_D^p & \text{if } T_n \geq 3h \end{cases}$$

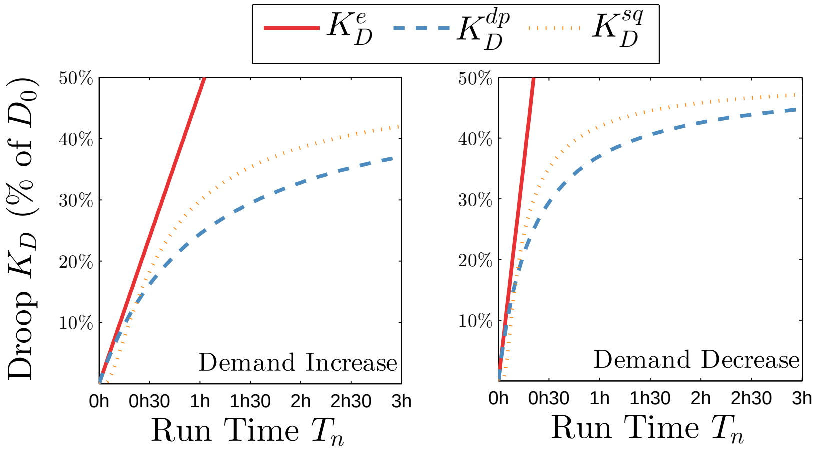


Figure 7.17: The different droop bounds K_D for signal tracking with bounded energy content $c(a = 3h)$ and $\Delta p = 50\%$. The most constrainting limit is K_D^{dp}.

7.5 Inverse model solving in autonomous control

In the autonomous control context, the information accessible to each load is limited to the load's parameters and the reference output $r^e(t) = (f(t) - f_n)/\phi$. In order to determine at which power rate $p(t)$ it should start running, a load must however rely on some information about the past input $p(t)$ that would have been necessary to follow this reference signal. Indeed, it is required to define the value of $\tau(t)$ and inverse the equation system.

Let's suppose a load desires to know the input $p(t)$ that would allow its group to follow respond to the frequency error without rebound error. It starts observing the frequency from $t = 0$ and needs to initialize the inverse model that must be solved. Fortunately, thanks to the filtering process, it is possible to find an initial condition that will asymptotically lead to the correct solution, after some learning period.

Proposition 5. *The input value $p(t, \nu)$ at time $t > 0$ that is the solution of the inverse model $y(t \geq 0) = T_n K_D r^e(t)$ with known initial values $p(t) = \nu(t), \forall t \in [-\infty, 0)$ are asymptotically convergent for all $\nu(t)$ iff these lead the group to have stored the same amount of energy $E_s(0)$ at time $t = 0$.*

Proof. Let's suppose that their exist two solutions $p_1(t)$ and $p_2(t)$ relative to different initial conditions $\nu_1(t)$ and $\nu_2(t)$.

The first side of the proof is trivial. If two functions must stay equal at all time $t > 0$ $p_1(t) = p_2(t) = p(t)$, and that they lead to the same output $y(t) = p(t)\tau_1(t) = p(t)\tau_2(t)$, then the run time is unique $\tau_1(t) = \tau_2(t) = \tau(t)$ at all time $t > 0$. Recalling that the run time evolution is defined by its time derivative $\dot{\tau}(t) = 1 - \frac{p(t)}{p_1(t-\tau(t))} = 1 - \frac{p(t)}{p_2(t-\tau(t))}$, this suggests that $p_1(s) = p_2(s) \forall s \in [-\tau(t), t]$. Consequently, $\nu_1(t) = \nu_2(t)$.

The second side is a little more complex. We would like to illustrate the following developments by introducing the remaining time function $T_r(t,s)$.

$$T_r(t,s) = \begin{cases} T_n & \forall t, s \geq 0 \qquad (7.62) \\ T_n - \displaystyle\int_{t+s}^{t} p(z)dz & \forall t, s \in [-\tau(t), 0] \qquad (7.63) \\ 0 & \forall t, s \leq -\tau(t) \qquad (7.64) \end{cases}$$

This function is a two variable function that represents the time remaining to a load before it reaches the end of its run time. This time is evaluated at time t for all loads which arrival occurred at time $t+s$. The function is well defined for $s \in [-\tau(t), 0]$. Therefore, we impose $T_r(t,s) = T_n$ for all $s \geq 0$ (load has not been running yet) and $T_r(t,s) = 0$ for $s \leq \tau(t)$ by definition of the dynamic run time $\tau(t)$.

The partial derivatives of $T_r(t,s)$ w.r.t. t and s are linked in the interval $s \in [-\tau(t), 0]$.

$$\frac{\partial T_r(t,s)}{\partial t} = -\frac{\partial}{\partial t}\int_{t+s}^{t} p(z)\,dz \qquad (7.65)$$

$$= p(t+s) - p(t) \qquad (7.66)$$

$$\frac{\partial T_r(t,s)}{\partial s} = -\frac{\partial}{\partial s}\int_{t+s}^{t} p(z)\,dz \qquad (7.67)$$

$$= p(t+s) \qquad (7.68)$$

This leads to the following expression, for all positive time t (i.e., where $y(t) = p(t)\tau(t)$) and $s \in [-\tau(t), 0]$.

$$\frac{\partial T_r(t,s)}{\partial s} - \frac{\partial T_r(t,s)}{\partial t} = p(t) = \frac{y(t)}{\tau(t)} \qquad (7.69)$$

We denote by respectively $T_{r,1}(t,s)$ and $T_{r,2}(t,s)$ the remaining run time of the two groups with power input $p_1(t)$ and $p_2(t)$. An illustrative view on this is shown on figure 7.18. The area under both curves is the remaining energy of the two group $E_r(t)$ and is the same for both of these inputs as from $t = 0$. Therefore, we have the following.

$$\int_{-\tau_1(t)}^{0} T_{r,1}(t,s)\, ds = E_r(t) = \int_{-\tau_2(t)}^{0} T_{r,2}(t,s)\, ds \qquad \forall t \geq 0 \quad (7.70)$$

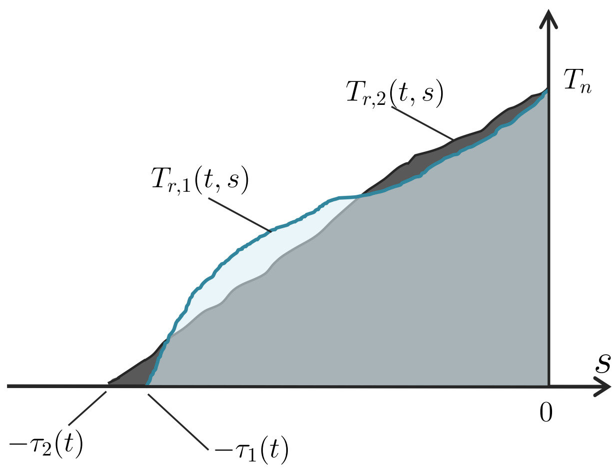


Figure 7.18: Remaining run time functions of input $p_1(t)$ and $p_2(t)$. The area under the curve is the remaining energy $E_r(t)$ of the group. It corresponds to the sum of the stored energy $E_s(t)$ with the natural value of the remaining energy $E_{r,n} = T_n^2/2$.

The time derivative of the remaining energy is the difference $T_n - y(t)$, that we can equal to the time derivative of both integrals.

$$\frac{d}{dt}\int_{-\tau_1(t)}^{0} T_{r,1}(t,s)\, ds = T_n - y(t) = \frac{d}{dt}\int_{-\tau_2(t)}^{0} T_{r,2}(t,s)\, ds \qquad \forall t \geq 0 \, (7.71)$$

Let's define the function $V(t)$ that is the squared difference of the area between these two curves. This function takes two different forms, depending on the relative value between $\tau_1(t)$ and $\tau_2(t)$.

$$V(t) = \int_{\max(-\tau_2(t),-\tau_1(t))}^{0} \big(T_{r,1}(t,s) - T_{r,2}(t,s)\big)^2\, ds \qquad t \geq 0 \quad (7.72)$$

Let's suppose that $\tau_2(t) > \tau_1(t)$. We then apply the Leibniz rule to derive $V(t)$ w.r.t. time t.

$$\frac{dV(t)}{dt} = \int_{-\tau_1(t)}^{0} \frac{\partial}{\partial t}T_{r,1}^2(t,s) + \frac{\partial}{\partial t}T_{r,2}^2(t,s) - 2\frac{\partial}{\partial t}\big(T_{r,1}(t,s)T_{r,2}(t,s)\big)\, ds \quad (7.73)$$

Following the Leibniz integration rule, we have considered the time derivatives of the boundaries $\frac{d\tau_1(t)}{dt}$ and $\frac{d\tau_2(t)}{dt}$, but their contribution is zero, as they multiply respectively $T_{r,1}(t, s \leq -\tau_1(t)) = 0$ and $T_{r,2}(t, s = -\tau_2(t)) = 0$. We can know develop.

$$\begin{aligned}
\frac{1}{2}\frac{dV(t)}{dt} &= \int_{-\tau_1(t)}^{0} T_{r,1}(t,s)\frac{\partial}{\partial t}T_{r,1}(t,s) + T_{r,2}(t,s)\frac{\partial}{\partial t}T_{r,2}(t,s) \\
&\quad - \big(T_{r,2}(t,s)\frac{\partial}{\partial t}T_{r,1}(t,s) + T_{r,1}(t,s)\frac{\partial}{\partial t}T_{r,2}(t,s)\big)\, ds \qquad (7.74)\\
&= \int_{-\tau_1(t)}^{0} \Big(T_{r,1}(t,s) - T_{r,2}(t,s)\Big)\frac{\partial}{\partial t}\Big(T_{r,1}(t,s) - T_{r,2}(t,s)\Big)\, ds \qquad (7.75)
\end{aligned}$$

Then, let's substitute the partial derivative of $T_r(t,s)$ w.r.t. t by the one w.r.t. s thanks to (7.69).

$$\begin{aligned}
\frac{1}{2}\frac{dV(t)}{dt} &= \int_{-\tau_1(t)}^{0} \Big(T_{r,1}(t,s) - T_{r,2}(t,s)\Big)\frac{\partial}{\partial s}\Big(T_{r,1}(t,s) - T_{r,2}(t,s)\Big)\, ds \\
&\quad - \int_{-\tau_1(t)}^{0} \Big(T_{r,1}(t,s) - T_{r,2}(t,s)\Big)\left(\frac{y(t)}{\tau_1(t)} - \frac{y(t)}{\tau_2(t)}\right) ds \qquad (7.76)\\
&= \frac{1}{2}\Big[\big(T_{r,1}(t,s) - T_{r,2}(t,s)\big)^2\Big]_{-\tau_1(t)}^{0} \\
&\quad - \left(\frac{y(t)}{\tau_1(t)} - \frac{y(t)}{\tau_2(t)}\right)\int_{-\tau_1(t)}^{0}\Big(T_{r,1}(t,s) - T_{r,2}(t,s)\Big)\, ds \qquad (7.77)\\
&= \frac{1}{2}(T_n - T_n)^2 - \frac{1}{2}\big(0 - T_{r,2}(t,-\tau_1(t))\big)^2 \\
&\quad - \left(\frac{y(t)}{\tau_1(t)} - \frac{y(t)}{\tau_2(t)}\right)\left(E_r(t) - (E_r(t) - \int_{-\tau_2(t)}^{-\tau_1(t)} T_{r,2}(t,s)\, ds)\right) \qquad (7.78)
\end{aligned}$$

This leads to the following.

$$\frac{dV(t)}{dt} = -T_{r,2}^2(t,-\tau_1(t)) - 2y(t)\left(\frac{1}{\tau_1(t)} - \frac{1}{\tau_2(t)}\right)\int_{-\tau_2(t)}^{-\tau_1(t)} T_{r,2}(t,s)\, ds \qquad (7.79)$$

Recalling that $y(t) > 0$ and $T_{r,2}(t,s) > 0$, the time derivative of $V(t)$ is negative in the case $\tau_2(t) > \tau_1(t)$. As this is completely symmetric, it is also the case for $\tau_1(t) > \tau_2(t)$ (adapting all integral bounds). Furthermore, it will be zero only when $\tau_1(t) = \tau_2(t)$ permanently. We have found a positive function V(t) whose time derivative is always negative except if $\tau_1(t) = \tau_2(t)$, and thereby $p_1(t) = p_2(t)$. This mean that all functions $p_1(t)$ and $p_2(t)$ are asymptotically convergent (as V(t) converges, and when it does, it can only converge to zero). $\square$

Autonomous computation of the Stored Energy Level

With the filtering process that the reference signals goes through, the stored energy can be computed from a limited amount of information.

$$E_s(t) = K_D T_n \int_{-\infty}^{t} r^e(s)\, ds = K_D T_n \int_{t-a}^{t} r^e(s)\, ds \tag{7.80}$$

The relevant information is indeed situated in the limited interval $[t - a, t]$. This has a twofold advantage in terms of practical implementation. Firstly, the necessary memory required to store the relevant information can be drastically limited. Secondly, a load that gets connected to the grid for the first time could determine remotely, after a *filtering period* of duration a, the actual stored energy of its group without any communication. Then a *learning period* is necessary for the inverse model to converge to a sufficiently accurate input. This is also valid for a load that was completely disconnected from the grid for any reason (restoration after blackout, etc.).

As shown by proposition 5, the inverse model can be initiated with any initial condition $\nu(t)$ that results in a stored energy equal to the one obtained from filtering the frequency with adequate parameter. Altogether, there are two conditions defining the suited initial conditions.

$$\int_{-\tau_0}^{0} \int_{-\tau(0)}^{\theta} \nu(s)\, ds d\theta = \frac{T_n^2}{2} + K_D T_n \int_{t-a}^{t} r^e(s)\, ds \tag{7.81}$$

$$\int_{-\tau_0}^{t} \nu(s)\, ds = T_n \tag{7.82}$$

The most simple choice consists in supposing a constant initial input ν_e as if the system had reached its steady-state corresponding to the same stored energy (prop. 4).

$$E_r(0) = \frac{T_n^2}{2} + K_D T_n \int_{t-a}^{t} r^e(s)\, ds \tag{7.83}$$

$$\nu_e = \frac{T_n^2}{E_r(0)} \qquad \tau(0) = \frac{2E_r(0)}{T_n} \tag{7.84}$$

The resulting inputs $\hat{p}(t, \nu)$ are estimates of the exact $p(t)$. We can test the convergence rate of the estimation by simulation. A system with loads of duration T_n starts from steady-state and follows the frequency signal. New loads get connected and must both filter the frequency, compute the resulting stored energy after the filtering period and then start estimating $\hat{p}(t, \nu_e(t_1))$, where t_1 is the time at which the state estimation starts. We will be able to compare the real value of $p(t)$ to their estimations.

Some results are shown on figure 7.19. The convergence time ($= 1.5T_n$) is defined as the time after which the following estimation error e_{SE} is below 0.1%.

$$e_{SE}(t, t_1) = \frac{\hat{p}(t, \nu_e(t_1))}{p(t)} - 1 \tag{7.85}$$

The information that a load needs prior to its start in order to perform its frequency response with high accuracy is equal to the sum of the filtering period a (i.e. to determine the energy storage level) and the learning period $1.5T_n$ required to reach sufficient tracking accuracy.

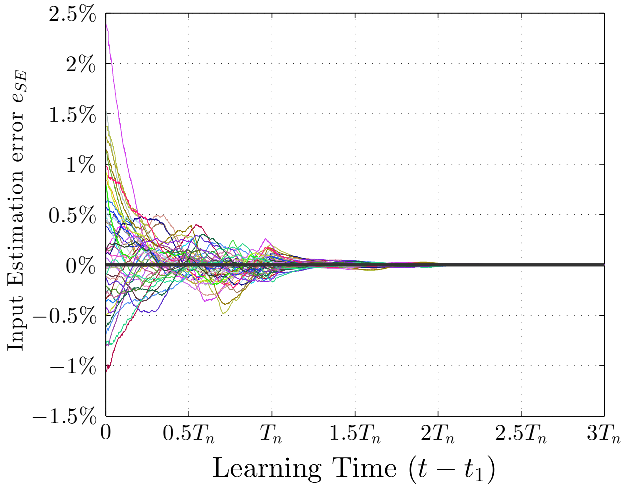


Figure 7.19: Convergence of the inputs in state estimation (learning). The tests where conducted with a filtering period $a = 3h$ on loads with $T_n = 1h$. Altogether, the required memory spans on a period $1.5T_n + a = 4.5h$.

7.6 Individual Load Simulation: Setting

The objective of this section is to compare the advanced power control performance to the proportional framework explored in the previous chapter. To this end, we conduct a detailed simulation in which each load is separately modeled. Our goal is to represent what would be the consequences of completely replacing the full generation-based FCR_G by demand-side control.

For simplicity, we focus on a single scenario which corresponds to participating loads having random power rate $P \in [2,7]kW$ and run time $T \in [1/4h, 3h]$ that are uniformly distributed. We suppose that loads can set their power freely in the interval $[p_L, p_H] = [50\%, 150\%]P$ (rate control policy).

Individual Power Controller in heterogeneous groups

We compare the impact of three different local power controllers. We refer to these three controllers by subscripts $\{1,2,3\}$ below.

1. *Advanced Power Control.* All elements introduced in this chapter are implemented.

2. *Proportional frequency response.* This is similar to what was presented in details in previous chapters (Rate policy).

3. *Proportional response with frequency filtering.* An intermediate set-point definition in which loads filter the system frequency for energy management purpose, and then respond proportionally to the filtered frequency.

The local controllers must all be designed in the same way. In our simulations, the frequency filter is chosen with a unique parameter $a = 3h$. Yet, the output (power set-point) will be relative to local load parameters. In particular, the inverse model and capacity allocation processes have results that are dependent on the load's random run time T. There exist two main ways to to manage this parameter dependency.

a. *Top-down approach.* Power set-points can be defined based on aggregation models that fully capture the heterogeneous nature of the group in which the load lies. This requires exploiting the parameter distribution, which is however not easily accessible at local level.

b. *Bottom-up approach.* A load reacts as if it was in an homogeneous group. It will consider that all other running loads share the same run time than itself. If the group is sufficiently large, the different contributions of the subgroups sharing identical run time will sum up smoothly. Furthermore, all the necessary information can be guessed from local measurements or knowledge.

As our goal is to design a control framework that would work as autonomously as possible, we will consider the second approach (bottom-up).

ACE definition and Aggregation model

It is expected that the integral action from aFRR will play an important role for rebound compensation. The definition of the ACE, which serves as basis for the centralized aFRR set-point definition, has very large consequences on the effectiveness of the load's contributions to overall system stability. We will consider two ACE definitions (1) Ideal ACE^{id} based on perfect information and group's demand measurement $D(t)$ and (2) Theoretical ACE^{th} based on aggregation models. Both are defined below, and are basically a scaled version of the group's rebound. The traditional ACE definition (based on the actual frequency deviation) is explored in the event-based simulation.

The ideal and theoretical ACE must be adapted according to the three strategies (1,2 and 3) at stake. Indeed, the expected response of the group depends on whether a capacity allocation process has intentionally reduced the offered droop K_D, which technical limit is $K_D^p = 1-p_L = p_H-1 = 50\%$. In what follows, the size of the group is chosen such that the loads fully replace the FCR_G in the advanced power control framework, with a reduced $K_D^{dp} = 32\%$ (more loads are required). Each case is simulated independently (different dynamics). Even though it make the text more cumbersome, we add for clarity a subscript (i.e., 1,2 and 3) relative to each case as well as a superscript referring to the ACE definition type.

$$ACE_1^{id}(t) = D_1^{id}(t) - D_0(1 + K_D^{dp}\frac{\Delta f_1^{id}(t)}{\phi}) \tag{7.86}$$

$$ACE_{2,3}^{id}(t) = D_{2,3}^{id}(t) - D_0(1 + K_D^{p}\frac{\Delta f_{2,3}^{id}(t)}{\phi}) \tag{7.87}$$

With regards to the theoretical ACE, the inverse model and capacity allocation insure that the provided FCR capacity K_D^{dp} in the advance control framework behaves as if all loads had a run time equal to the frequency filter parameter $a = 3h$. The rebound consists therefore in a scaled version the frequency offset imposed to the system frequency.

$$ACE_1^{th}(t) = -K_D^{dp}D_0\frac{1}{a}\int_{t-a}^{t}\frac{\Delta f_1^{th}(s)}{\phi}\,ds = K_D^{dp}D_0\frac{\Delta f_1^{e,th}(t) - \Delta f_1^{th}(t)}{\phi}. \tag{7.88}$$

In the two other cases (proportional response with and without filtering), the rebound is the natural proportional response of the group to respectively the filtered and non-filtered frequency deviations. Let's also note that both will be simulated independently. In the equations below, we denote by $T_n = 0.5(T_M + T_m)$ the average run time at start.

$$ACE_2^{th}(t) = K_D^p D_0 \left[\frac{\Delta f_2^{e,th}(t) - \Delta f_2^{th}(t)}{\phi} - \frac{1}{T_n} \int_{T_m}^{T_M} \int_{t-T}^{t} \frac{\Delta f_2^{e,th}(s)}{\phi} \, ds \Omega_T \, dT \right] \quad (7.89)$$

$$ACE_3^{th}(t) = -K_D^p D_0 \frac{1}{T_n} \int_{T_m}^{T_M} \left(\frac{1}{T} \int_{t-T}^{t} \frac{\Delta f_3^{th}(s)}{\phi} \, ds \right) T \Omega_T \, dT \quad (7.90)$$

Individual load simulation and frequency measurements errors

Our goal is to simulate each individual load arriving either at random or fixed rate λ and equipped with advanced power controller and frequency sensors. Individual simulations are complex, and require a lot of memory and computation.

In the simulation, only the necessary information is carried out within the different computation blocks (i.e. inverse model, frequency filtering). The inverse model is solved with an efficient discrete time approach. Yet, it was not possible to simulate all individual loads that would be necessary in practice to cover the FCR_G requirements. A scaling factor of about 100 was necessary, and the simulation length was limited to 3 days.

This is however sufficient to compare the different control frameworks, and their potential impact on the system as a whole. One of the interesting feature of the individual loads simulation is that all above discussed elements can be tested together with local considerations.

- *Frequency measurements errors.* The frequency sensors must have very limited cost. Therefore, they could be subject to measurement errors. The frequency measurement error at load i has in general a time varying component (noise) $\epsilon_i(t)$ that is zero mean and of variance σ_e^2 as well as a constructive bias f_i^b of known distribution Ω_{f^b}. In what follows, we will test the impact of individually errors on the control performances.

- *Disturbance rejection through frequency filtering.* Thanks to the filtering procedure, frequency errors cannot accumulate indefinitely. Indeed, the noisy frequency integral is limited to period a (filter parameter). Any fixed bias in the measurement is fully rejected from the instantaneous value of the filtered frequency after the filtering period a. It disappears from the integral of the filtered signal after another period a. The standard deviation of the noise $\epsilon_i(t)$ has also an impact on the measured energy state. Indeed, the time integral of a zero mean white noise is zero mean but its standard deviation increases with the square-root of the time period on which the integral is performed.

Altogether, the group's energy error, when averaged out among all running loads is approximately equal to $\sigma_e\sqrt{a/N_0}$, where N_0 is the number of running loads at rest. For loads that are not equipped with the filtering block (third simulated case), this noise impact will reach $\sigma_e\sqrt{T_n/N_0}$ while the bias will lead to a permanent control error proportional to $T_n\mathbb{E}[f^b] = T_n \int f^b \Omega_{f^b}\, df^b$.

- *Inverse model with state estimation.* After the filtering period, a learning period starts. Each load will need to estimate the run time $\tau(t,T)$ of the hypothetical homogeneous group they are part of. The convergence is guaranteed, while the convergence time is evaluated from simulation to be about $1.5T$. In our simulations, the load enters the system without any knowledge about the past. For the group to reach rapidly its initial equilibrium (i.e., $D(t) = D_0$), we impose a parameter independent learning period for each load equal to $1.5T_M$.

Let's note that with the ideal ACE definition, part of the measurement errors will be compensated by aFRR. It would also be the case in practice, as the impact of such errors on the system frequency would eventually need to be compensated by an integral action.

System scaling and impact on Demand's volatility

As it soon becomes impractical to simulate all loads required for delivering the $3GW$ of FCR, we will need to scale some quantities. In practice, we consider loads to be between 10 and 20 times bigger $P \in [20, 30]$ and the system to be 10 times smaller ($FCR_D = 300MW$). Such approximation lead's the group's demand to be much more variable than what would be observed in practice.

The parameter λ_{sim} will be situated between 4 and 2 in the simulations while it would be 100 to 200 times larger in real-life $\lambda = 100\lambda_{sim}$ or $\lambda = 200\lambda_{sim}$. In real-life, the relative parameters variances would respectively be $(\sigma_T^r)^2 = \frac{2(10.8-0.9)^2}{12(10.8+0.9)}10^3 = 1.4\,10^3$ and $(\sigma_P^r)^2 = \frac{2((7-2)*1e-3)^2}{12(2+7)*1e-3} = 4.6\,10^{-4}$. In the simulation, the power rate variance is 10 to 20 times bigger. The relative volatility of the group in the simulation $v_{D,sim}$ is derived below and compared to its real-life value v_D.

$$v_{D,sim10} = \frac{\sqrt{(10\sigma_P^r)^2+1}}{\sqrt{0.01\lambda T_n}} = 0.7\% = 10v_D$$

$$v_{D,sim20} = \frac{\sqrt{(20\sigma_P^r)^2+1}}{\sqrt{0.005\lambda T_n}} = 1\% = 14.2v_D$$

We can therefore expect that the baseline in the simulation will varying of $\pm 3v_{D,sim20} = 3\%$ around its expected value to parameter and starting rate variability. It would have been 0.2% if we were able to simulate at correct scale. This is a problem, knowing that the percentage of the baseline demand that can actually be offered as FCR is limited by the capacity allocation process to values that get dangerously close to the volatility interval.

We recall for clarity that P is random, and that P_n here expresses the expected value of P for the whole group. Let's suppose that loads can set their power freely in the interval $p(t) =\in [-\Delta p, \Delta p]$ relatively to their nominal power P. The capacity allocation process with a filtered frequency of parameter $a = 3h$ is illustrated on figure 7.20 for $\Delta p = 50\%$ and $\Delta p = 30\%$. On this figure, the natural volatility interval $\pm 3v_{D,sim}$ of the baseline demand is also highlighted.

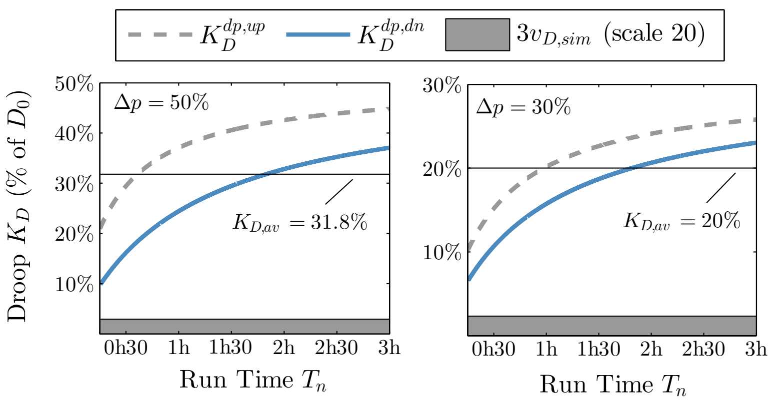


Figure 7.20: Derating of the offered capacity (i.e. controller's gain $K_D^{dp,dn}$ and $K_D^{dp,up}$) as computed by the capacity allocation mechanism, in function of the load's run time. Left: $K_D^p = \Delta p = 50\%$. Right: $K_D^p = \Delta p = 30\%$.

The average symmetrical droop $K_{D,av}$ that is offered by the group is dependent on the run time distribution $\Omega_T = (T_M - T_m)^{-1}$, of expected value $\mathbb{E}[T] = 0.5(T_m + T_M) = T_n$.

$$K_{D,av} = \frac{1}{\mathbb{E}[T]} \int_{T_m}^{T_M} K_D^{dp,dn}(T) T \Omega_T \, dT \tag{7.91}$$

Figure 7.21 shows the evolution of the ratio $3v_D/K_{D,av}$ for the considered group in function of the load's technical limit Δp.

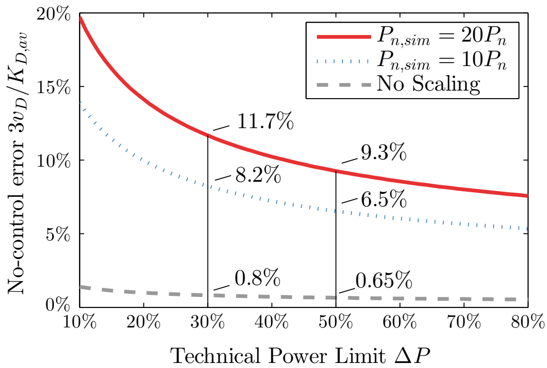


Figure 7.21: Maximal absolute control error at rest (i.e., $r^e = 0$) expected from the group depending on its loads' technical capability and on the necessary scaling factor.

Its seems necessary for analyzing the simulation results to allow a larger tolerance on the group's performance. This larger tolerance would however not be required in real-life implementation as the natural variability would lead to a maximum control at rest of about $3v_D = 0.6\%$.

7.7 Historical simulations

The group's response is simulated in a typical context. We have selected three days in the 2015 year. The first day is an average day, the two others are those within which the frequency error had respectively the most extreme positive and negative energy content.

7.7.1 Power system model

$$M\frac{d\Delta f(t)}{dt} = I(t) - L\Delta f(t) + aFRR(t) - (D_{1,2,3}(t) - D_0) \quad (7.92)$$

$$\tau_g \frac{daFRR(t)}{dt} = aFRR^{set}(t) - aFRR(t) \quad (7.93)$$

$$aFRR^{set}(t) = -K_p\left(ACE^{id,th}_{1,2,3}(t) + \frac{1}{\tau_i}\int ACE^{id,th}_{1,2,3}(t)\,dt\right) \quad (7.94)$$

The exact value of $D(t)$, the aggregate group demand, is the result of the individual simulation. Our objective at this stage is to simulate every load

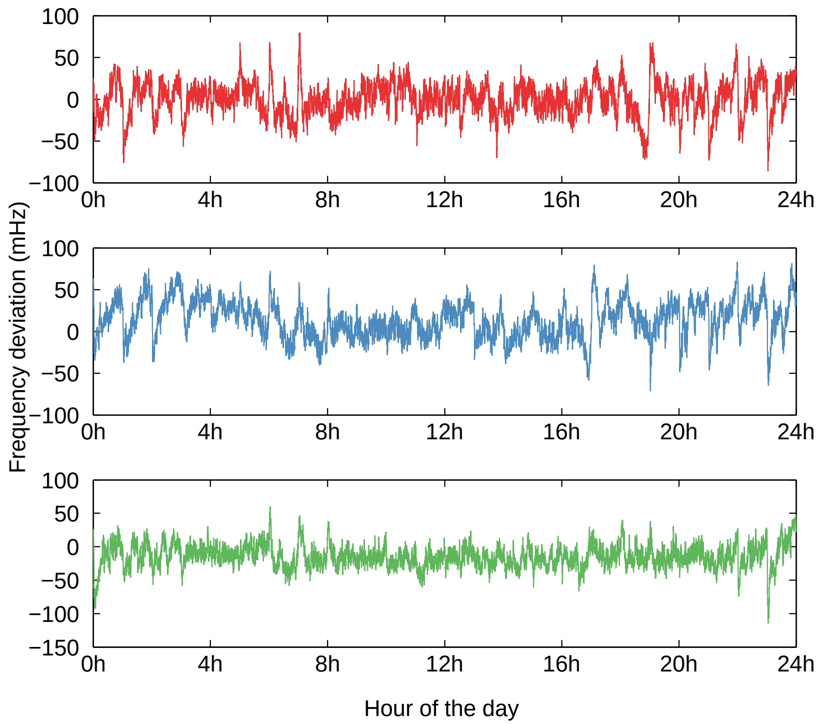


Figure 7.22: Frequency deviations along the three simulated days (source: RTE).

to evaluate the impact of the different source of errors: non-ideal frequency measurements, parameter distributions, random load arrivals. In long-run simulations, $D(t)$ is replaced by the aggregate model $\hat{x}(t)D_0$ to drastically limit computational burden.

Let's note that all above variables should be indexed by the control case (i.e., 1,2 and 3) and the ACE definition. For readability, we only emphasize by indexes and superscripts the elements which are defined specifically in the different cases. In this model, there are no FCR volumes provided by generation. All system quantities (e.g., imbalance $I(t)$, inertia M) are scaled by a factor 1/10 compared to their actual value. Power rates are scaled by a factor 20.

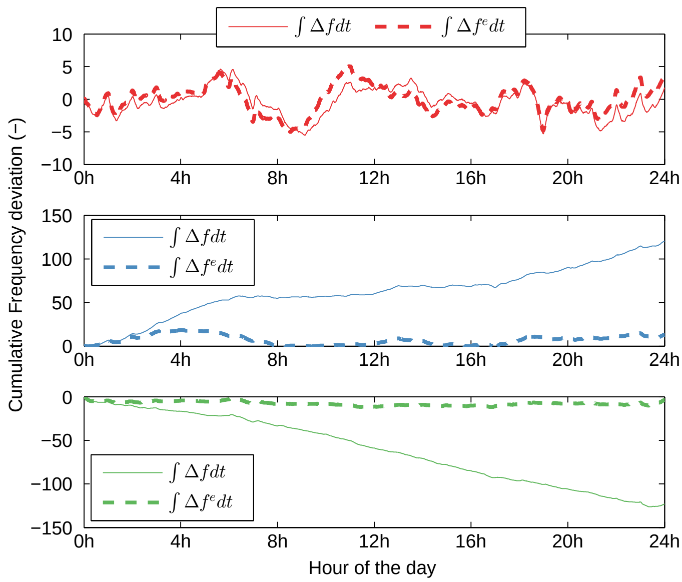


Figure 7.23: Evolution of the cumulative frequency deviations and its filtered couterpart along the three simulated days.

7.7.2 Results 1: Ideal ACE, random arrivals, ideal frequency measurements.

The main simulation results for all control cases and ideal ACE definition (direct group's demand measurement) are illustrated on figure 7.24.

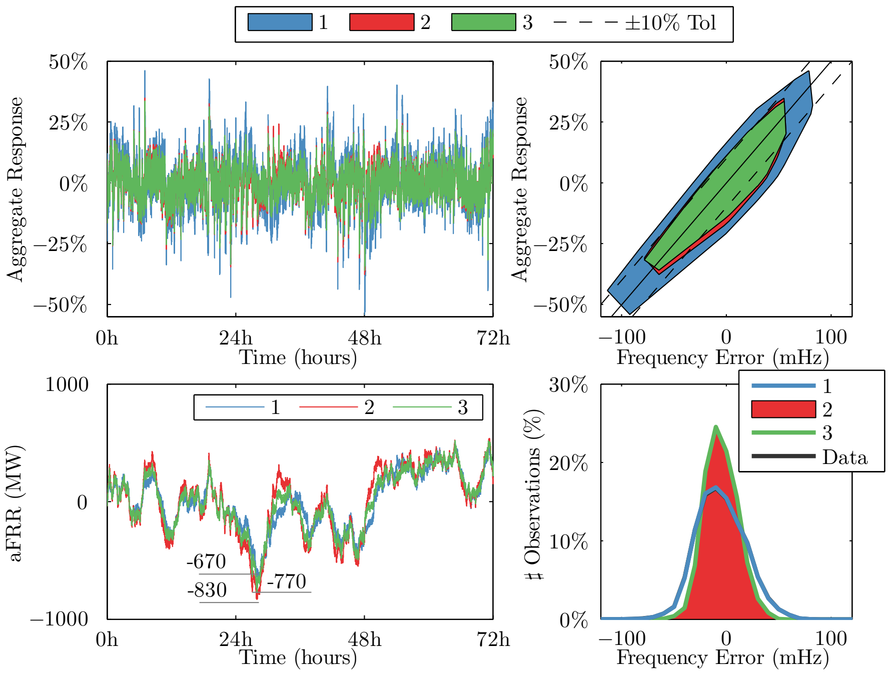


Figure 7.24: Simulation results for all control cases (1,2,3) and ideal ACE definition, random arrivals and ideal frequency measurements. Top left: Aggregate group response (in per unit). Top right: Convex hull of the group response in function of the frequency deviation. Bottom Left: additional aFRR volumes. Bottom Right: Frequency histogram and comparison with initial data.

First, and before reading all comments below, let's insist again on the methodology.

> In this simulation, **the number of loads is fixed** for all control cases. In addition, the group is **scaled** according to the advanced case (case 1), in order for the controlled loads in the advanced case to provide the **total FCR requirement** of in the CE system. This means that the proportional controller cases (2

and 3) have virtually more FCR at their disposal (no capacity allocation process). This is the reason that explains the overall better frequency quality observed in simulations relative to the second and third control cases.

According to our assumptions (i.e., uniform distribution Ω_T), the number of loads required to cover the full FCR needs would be 36% lower (i.e., Fig. 7.20, $= 1 - 31.8/50$) if the group was scaled according to case 2 and 3. The per load profitability in such case is therefore much higher. Provided that just the adequate number of loads would be equipped in practice with power controllers, this represents an important economic advantage for case 2 and 3.

To the contrary, the additional aFRR needs in case 2 and 3 lead to higher costs on the system side. As shown on the bottom right chart of figure 7.24, the larger needs are observed in the second case (-830 MW) where two sources of rebound accumulate. The filtering process induces an artificial *recovery* error while the group response adds a second *rebound* error on top of it, as it is not equipped to counteract its natural rebound.

Results of the aFRR rebound management

The summed contribution of the group's response together with the associated aFRR are shown on figure 7.25. As highlighted, the ideal rebound compensation from aFRR leads these summed responses to respond to frequency deviations with very high performances (i.e., maximum error of $\pm 2.5\%$ around the linear frequency set-point). This confirms the fact that the response error introduced by the parameter randomness is rather slowly evolving and can be almost perfectly be followed by slower reserves.

Improving the advanced control scheme

The advanced control case (case 1) may seem to have lower performances than the two other simple proportional controllers (case 2 and 3). Indeed, the frequency quality is lower (though still alike historical level). Due to the capacity allocation process, the advanced framework is only allowed to offer a reduced amount of flexibility to the grid than what is offered in the simple case. The group must indeed keep part of its flexibility for rebound management purpose. The group in case 1 exhibits a lower immediate responsiveness what decreases the frequency quality. This acts as a positive feedback (i.e. *snowball effect*): a lower group's responsiveness induce larger frequency deviations that lead to higher relative use of the offered capacity. Therefore, more energy must be shifted. The system dynamics accentuate this trend. Indeed, the faster/larger the FCR damping, the faster will the

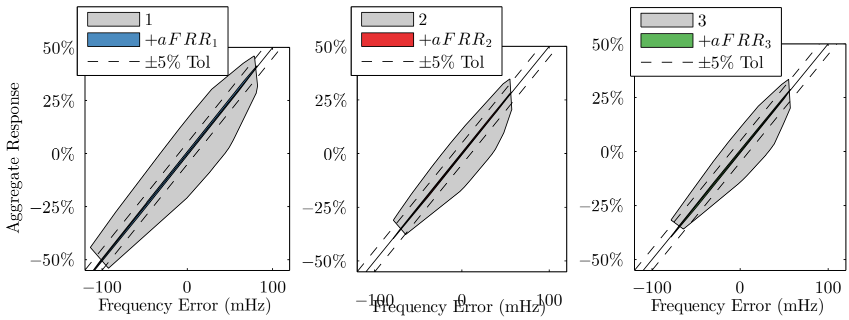


Figure 7.25: Convex hull of the group's response in function of the frequency deviation in each (i.e., 1,2,3) control case without (grey area) and with aFRR volumes (colored area).

system frequency reach steady-state, which limits the frequency deviation integral. In fact, the rotating masses lost less of their kinetic energy in the imbalance phase. The lower the frequency integral, the lower the required aFRR, as the they counteract the rebound/recovery error which is proportional to the frequency integral.

Yet, the advanced control scheme reduces the needs for aFRR. The additional requirements for downward reserve are 15% larger in the simple case (-770MW) than in the advanced case (-670MW) which raises the per load profitability of the advanced control. Altogether, the gains obtained on the aFRR side are smaller than the losses on offered capacity side. The proposed advanced controllers are preferable to their simple counterpart in systems with very high aFRR and FCR costs.

The low performances of the proposed advanced control scheme are the consequence the capacity allocation process. The capacity allocation process is overly restrictive because it takes a rather static view on the FCR deployment.

1. Firstly, a portion of the initial capacity must be kept for rebound management purpose, which limits is fixed by the energy bound $c(a)$.

2. In addition, the worst-case approach leads loads to restrict the offered capacity for risk management purpose. In fact, loads should always be deployed the offered capacity at all time, even in the worst case where they had just been being shifting the energy bound $c(a)$.

This static approach should be joined to a *dynamic deployment* of the capacity. The advanced control could be modified to allow loads to deploy *briefly* their full flexibility potential in order to rapidly contain frequency deviations in order to limit the integral of these deviations and therefore the rebound and their efforts to counteract it. Immediately after, the group would reduce progressively the deployed capacity as aFRR take over the burden. However, it is crucial that this is performed *in direct coordination* with aFRR. The set-point change should indeed be included in the ACE (e.g., computed externally by the TSO) rather than appear indirectly in the frequency deviation (otherwise, the integral of the frequency deviations would increase again!).

This process is rather complex and is unlikely to be widely accepted, as simple solutions are often preferred in practice

The ideal control scheme is therefore to be found between the simple and advanced scheme. In the simple scheme, the full flexibility potential ($= \pm \Delta p D_0$) is *initially* accessible. Yet, as soon as control begins, the rebound error degrades both the overall performances as well as the flexibility accessible in the future. Indeed, though inputs can still vary across the whole technical range $[p_L, p_H]$, the number of running loads evolves in opposition to past control actions. This lowers the amount of capacity that can be deployed by the group. If aFRR work sufficiently well, the technical impact on the grid is relatively small. In the advanced scheme, the initial flexibility potential is restricted by the *static* approach undertaken in the capacity allocation process.

A close look at the individual power inputs $p_i(t) \forall i \in N(t)$ observed in the simulations may corroborate this claim. They are presented on figure 7.26 together with the reference signal.

On the middle chart, the range of the simulated input relative to case 1 is shown. The range is rather large as the individual loads behave according to their parameters. The most extreme values of this range are relative to loads with the shortest run times. The inputs get closer from their acceptable limits p_L and p_H principally due to the efforts required to shift energy for rebound management. The remaining gap that separates observations from their limit is a consequence of the methodology and the selected worst-case approach that was used to derive the bounds on K_D^{dp}.

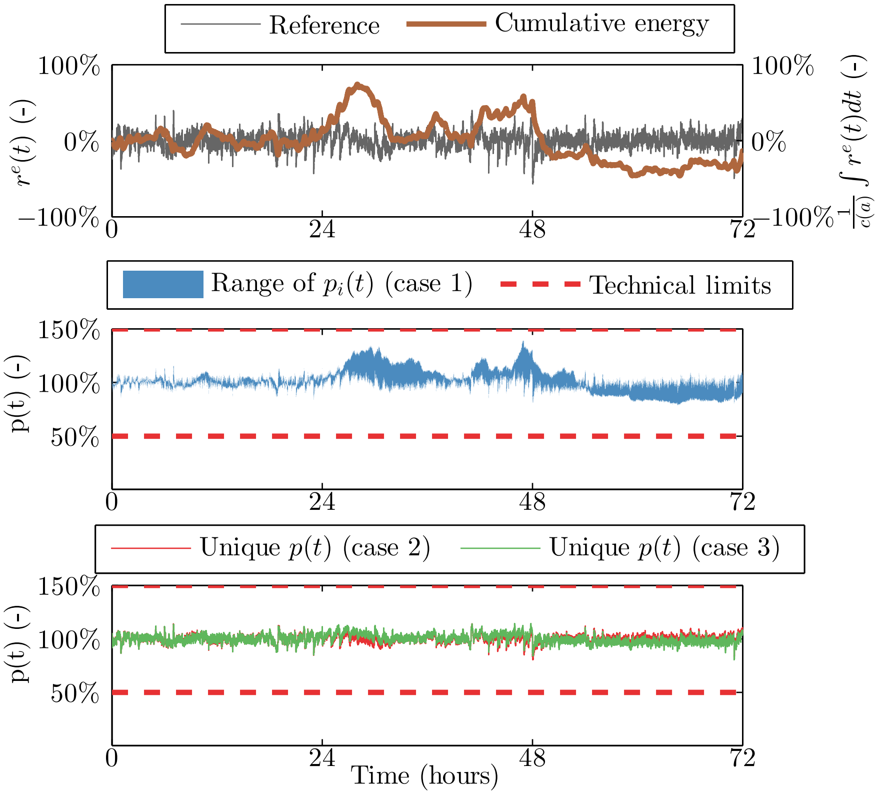


Figure 7.26: Range of the required local power inputs in the three control cases along the the 72 simulated hours (ideal ACE definition, random starts, no measurement error). In this simulation, input range originates from the load run time distribution. Top chart: reference signal $r^e(t)$. Middle chart: Range in the advanced control scheme (case 1). Bottom chart: Unique control input in the simpler schemes (cases 2 and 3).

7.7.3 Results 2: Ideal ACE definition, fixed arrival rate, measurement errors.

We now explore the impact of measurement errors on the simulation results, highlighted on figure 7.27.

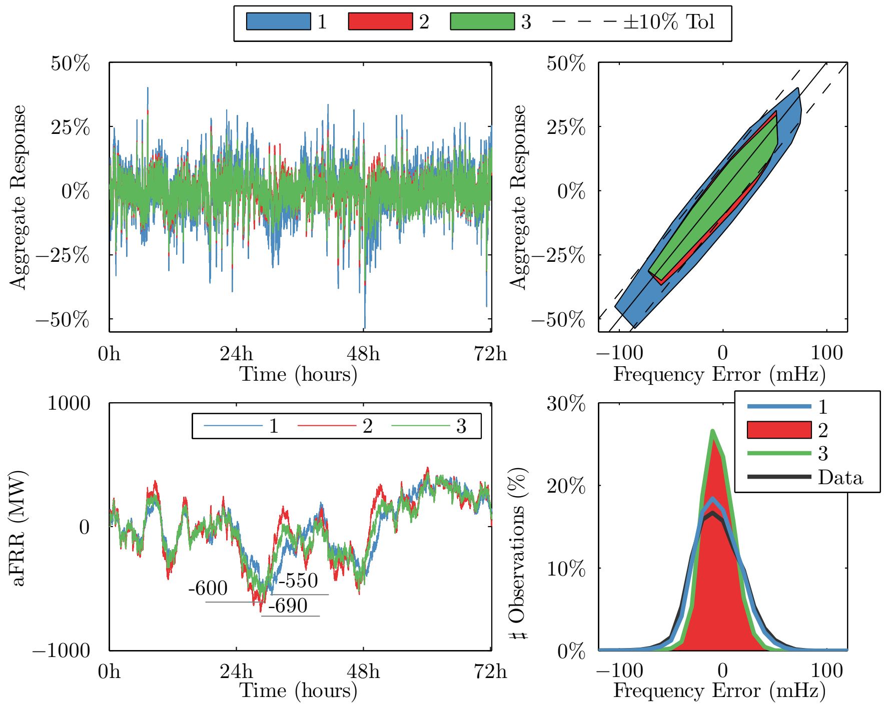


Figure 7.27: Simulation results for all control cases (1,2,3) and ideal ACE definition, constant arrivals and frequency measurement errors. Top left: Aggregate group response (in per unit). Top right: Convex hull of the group response in function of the frequency deviation. Bottom Left: additional aFRR volumes. Bottom Right: Frequency histogram and comparison with initial data.

A measurement bias of $f^b = 10$mHz is common to all loads (worst case) and measurements are subject to a random noise with standard deviation $\sigma_e = 1$mHz. On the system-side, results are qualitatively similar with or without considering these measurement errors. The aFRR requirements are a however a little increased as a consequence of the instantaneous errors, and the integration of the noise in the inverse model. In case 1, the minimum requirements are shown to be -550MW. Without measurement errors, the simulation (not shown) gave a minimum of -540MW, and the above results with random arrivals rate lead to -670MW. Measurement errors are very well rejected by the filtering process, as perceived from the group perspective.

However, errors have large consequences among the different loads. The constant measurement bias has no impact thanks to the filtering process. However, the rebound management process (inverse model) leads the loads to compensate for the integral of the noise term computed in the interval $[t-a,t]$. Virtually, loads tend to believe they need to shift an extra amount of energy which is normally distributed (integral of white noise) with zero mean and standard deviation $\sigma_e\sqrt{a}$. As highlighted on the middle and bottom charts of figure 7.28, the input ranges of the first and second cases are consequently larger than in the simulation without measurement error (fig. 7.26).

In addition, an input range appears in the second and third case due to the instantaneous error. It is relatively large as the main element of influence is the constructive bias. Indeed, in the simple case, the bias is not rejected as no filtering process is implemented. Furthermore, the bias was chosen to be ten times larger than the standard deviation of the noise. This means that the impact of the noise in case 1 should be approximately 30 times larger than the bias and noise in case 3 and 300 times the noise impact in case 2. The bias is rejected in case 2 by the filtering process, as visible from the bottom chart of figure 7.28: the range in case 2 is always below the one of case 3. This has however no impact on the response performance, as the baseline demand level is the same ($D_0 = \lambda E_n$, and energy needs E_n are conserved) with or without biased power.

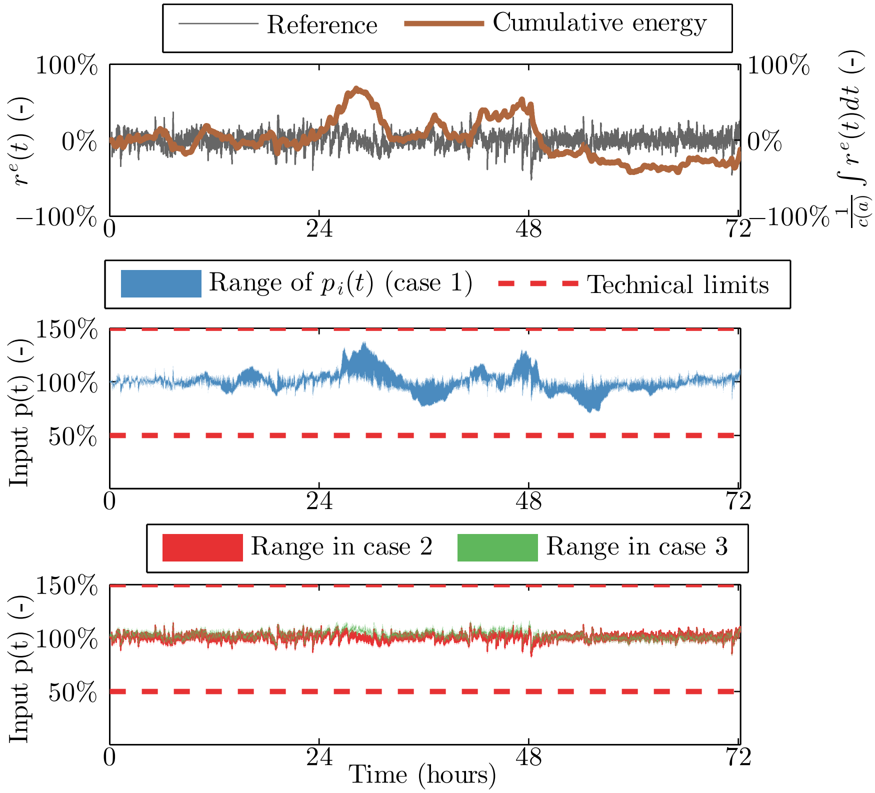


Figure 7.28: Range of the required local power inputs in the three control cases along the the 72 simulated hours with constant starts, ideal ACE definition and considering measurement errors. The input may differ from a load to the other due to different parameter (run time) or due to measurement error. Top chart: reference signal $r^e(t)$. Middle chart: Range in the advanced control scheme (case 1). Bottom chart: Ranges in the simpler schemes (cases 2 and 3).

7.7.4 Results 3: Theoretical ACE definition, random arrivals, measurement errors.

The same results in which the theoretical ACE definition is implemented are presented on figure 7.29: the ACE is defined only based on the frequency error without precise information about the actual group demand (aggregate model).

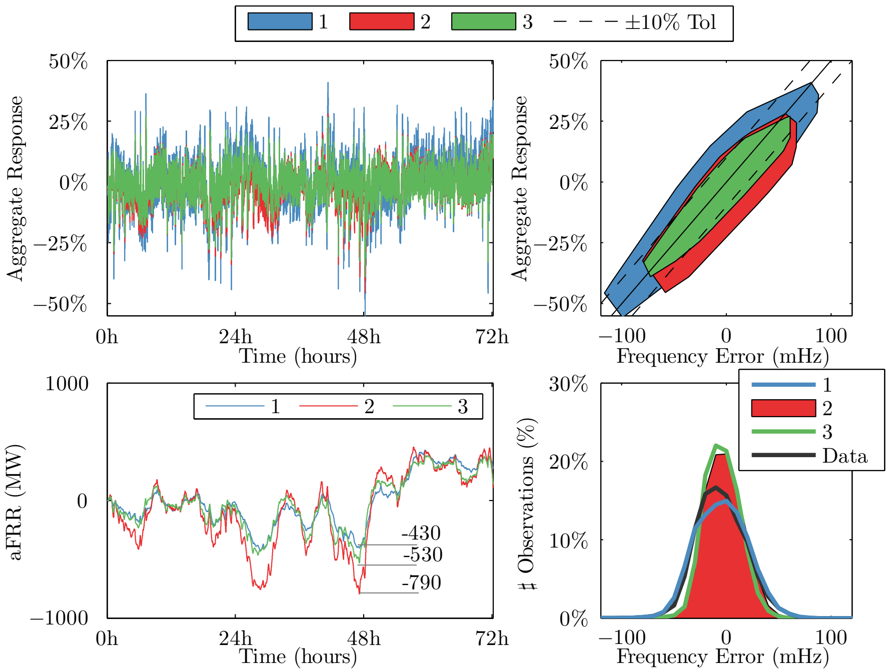


Figure 7.29: Simulation results for all control cases (1,2,3) and theoretical ACE definition, random arrivals and frequency measurement errors. Top left: Aggregate group response (in per unit). Top right: Convex hull of the group response in function of the frequency deviation. Bottom Left: additional aFRR volumes. Bottom Right: Frequency histogram and comparison with initial data.

For this centrally defined set-point to be precise enough and for guaranteeing system stability, the parameters distribution (i.e., natural run time and power rate, technical limits and starting rate) and therefore the group's baseline demand as well as the resulting offered FCR_D capacity should be known by the system operator. In the results below, loads start randomly with known starting rate and are subject to frequency measurement errors (fixed bias of $f^b = 10$mHz with random noise of std dev. $\sigma_e = 1$mHz).

The theoretical ACE is unable to capture all variable components of the actual demand $D(t)$. Yet, it performs relatively well, based on a limited amount of knowledge, as can be observed on figure 7.30. The exact realization of the random elements (exact number of arrivals, measurement errors) cannot be integrated *directly* in the ACE computation. However, they have a consequence on the frequency, which serves as basis in the theoretical ACE computation. This is what explains the good rebound compensation performance of aFRR in this case.

Yet, several elements should be observed.

- The maximum amount of deployed aFRR capacity is not observed at the same time with the ideal and theoretical ACE definition.
- The performances of the second control case 2 seems biased as observable from the convex hull presented on figure 7.29. However, it does not come from the measurement bias, which is rejected in the filtering process. The origins comes from the combination of recovery and rebound errors and that the integral of the simulated frequency is more often positive than negative. Each time the group's randomness leads to lower than expected reaction from the group, the theoretical ACE does not notice the change immediately. It has indeed to wait that the changes has an impact on the frequency. When the same phenomena occurs in the other direction (higher than expected reaction), the convex hull does not get larger, and such phenomena is therefore not observable in our results.

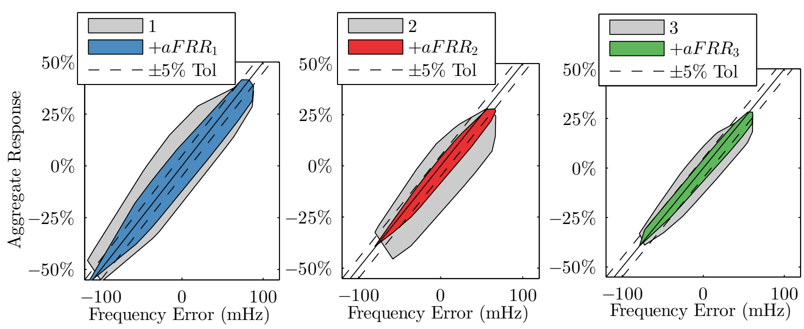


Figure 7.30: Convex hull of the group response in function of the frequency deviation with and without considering rebound management.

Consequences on loads input range

The individual power set-point must vary in a wider range, which gets dangerously close to the technical limits. This comes from the fact that the theoretical ACE does not directly captures the variation in the number of loads. The inputs must therefore compensates for it. In the ideal case, this variation was directly measurable.

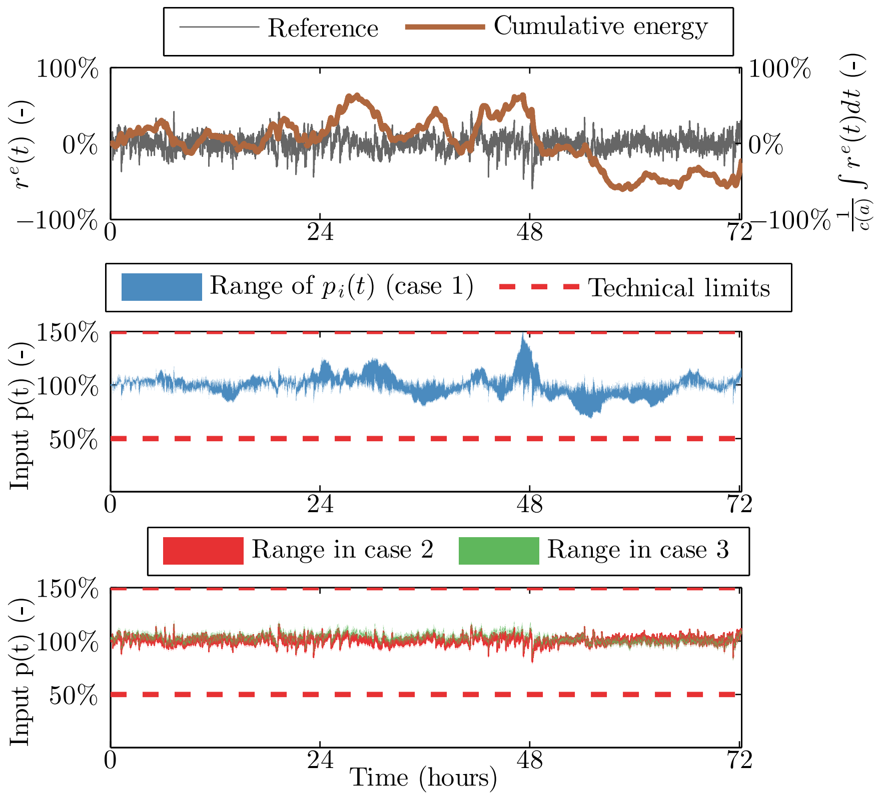


Figure 7.31: Range of the required local power inputs in the three control cases along the the 72 simulated hours theoretical ACE definition, with random starts and measurement errors. The input may differ from a load to the other due to different parameter (run time) or due to measurement error. Top chart: reference signal $r^e(t)$ and its time integral. Middle chart: Range in the advanced control scheme (case 1). Bottom chart: Ranges in the simpler schemes (cases 2 and 3).

7.7.5 Result 4: The effect of frequency filtering

For verification purpose, we conduct the same experiment but we omit the frequency filtering process. This will require virtually more energy to be shifted by the loads which will induce the local inputs to rapidly saturate. In addition, the biased and noisy measurement will get integrated in the inverse modeling which worsen the situation. This is shown on figure 7.32 where the effect of the measurement error is shown to induce input saturation. The circled portion of the range occurs is selected at a specific time in the simulation for which cumulative energy content of the non-filtered frequency is very small. This energy should not induce saturation, as loads are able to shift it thanks to the capacity allocation. This is, e.g., in opposition to what occurs at the 48^{th} hour of the simulation where saturation is mainly due to the high energy content of the non-filtered frequency.

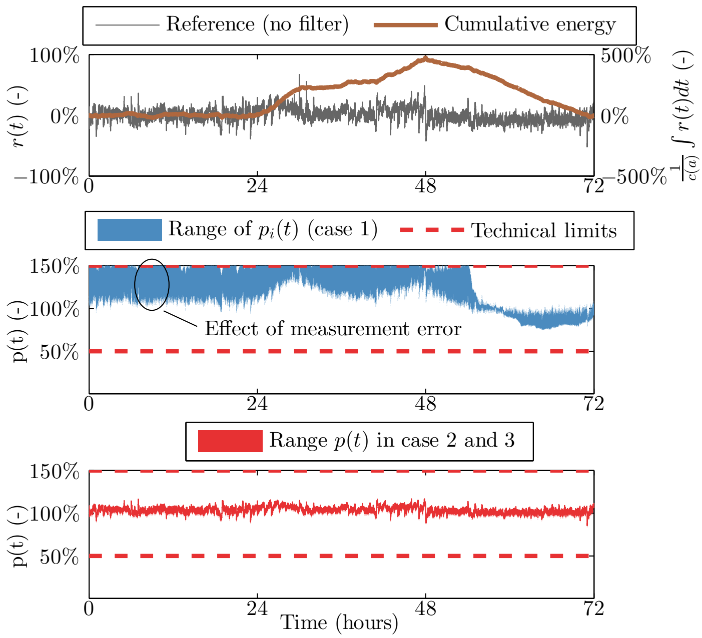


Figure 7.32: Range of the required local power inputs in the three control cases along the 72 simulated hours with ideal ACE definition, constant arrivals and measurement errors. Top chart: reference signal $r^e(t)$ and its time integral. Middle chart: Range in the advanced control scheme (case 1). Bottom chart: Unique input in control cases 2 and 3 (identical, as filter is omitted).

Due to the saturation, the tracking performances of the group degrade largely. On figure 7.33, the top left chart shows that the saturation impedes some loads to counteract frequency changes (i.e., downward regulation, positive frequency error) which lead the frequency to reach higher values than historically measured. In this case, aFRR were sufficiently fast to maintain system stability (small frequency changes). It would not have been the case if, e.g., a sudden contingency would had occurred.

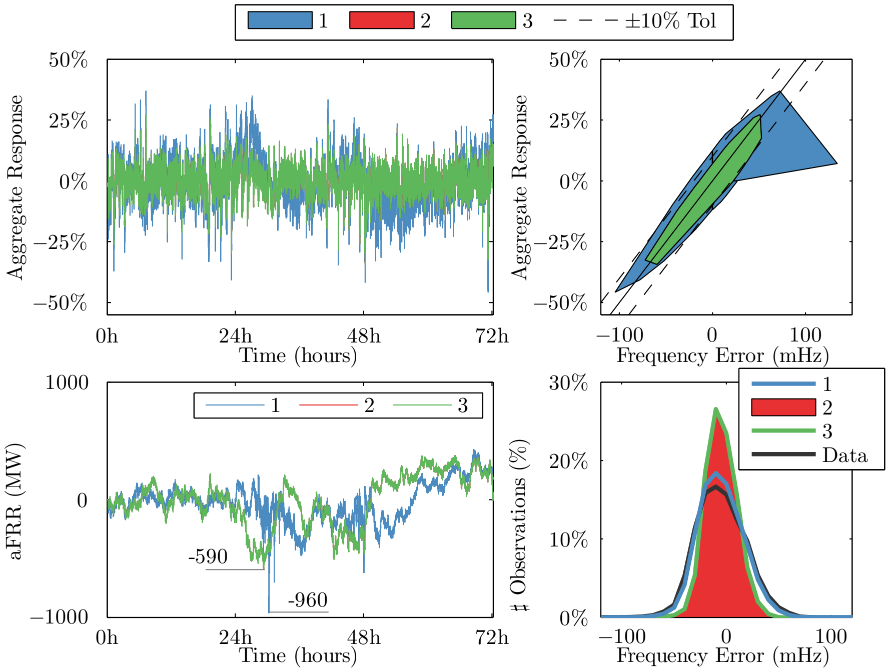


Figure 7.33: Simulation results for all control cases (1,2,3) without filtering process and measurement errors. Top left: Aggregate group response (relative to available capacity). Top right: Convex hull of the groups response in function of frequency deviation. Bottom Left: additional aFRR volumes (ideal rebound compensation). Bottom Right: Frequency histogram and comparison with initial data.

7.7.6 Event-based simulation, with random arrivals

We simulate the same control cases in which loads provide the full FCR volumes during a large contingency event: the sudden loss of 3000 MW in the EU power system. In this case, the ACE can be defined as the frequency error itself. Indeed, the system perturbation (imbalance $I(t)$) is well defined. The generation loss occurs at time $t = 30min$ and restored at time $t = 5h$. We can therefore observe the impact of a worst case scenario for both upward (event) and downward (restoration) regulation. Results are shown on fig. 7.34. The system is scaled to $1/10^{th}$ of its real size and loads are 20 time larger (power rate). The load arrive at random rate, which gives a more realistic flavor to the results.

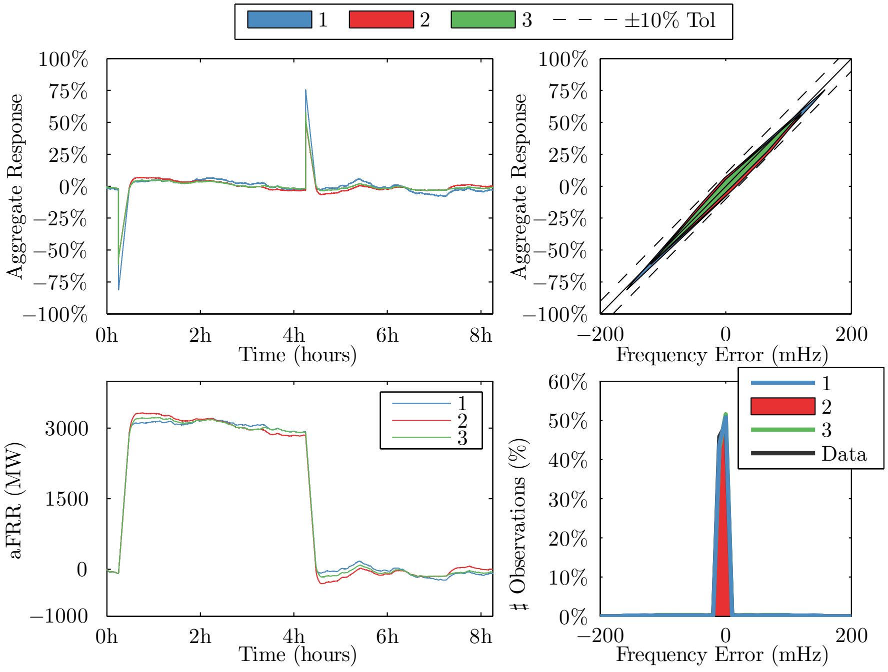


Figure 7.34: Results of the event-based simulations for all control cases (1,2,3) with random arrivals and ideal measurements. Top left: Aggregate group response. Top right: Convex hull of the groups response in function of frequency deviation. Bottom Left: additional aFRR volumes. Bottom Right: Frequency histogram and comparison with initial data.

The proportional response with frequency filtering (case 2) has again the less desirable behavior in terms of aFRR requirements. The advanced and simple control schemes differ slightly from each other in the amount of required aFRR but more strongly regarding the extreme frequency values. A closer look at the first hour of the simulation highlights these trends (Fig. 7.35).

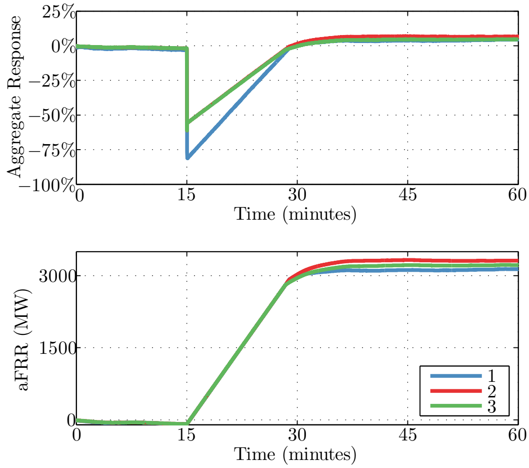


Figure 7.35: Detailed results in the first hour of the event-based simulations for all control cases (1,2,3) with random starts and ideal measurements. Top: Aggregate group response (relative to available capacity). Bottom: aFRR volumes (ACE is the frequency error).

The input range of all control cases is presented on figure 7.36. We can see that the capacity allocation process is strongly limiting the immediate response of small loads. This comes even clearer on the figure 7.37, where we zoom into the first event.

7.8 Aggregate model for long-term simulations

The individual simulations require a lot of memory and computational efforts. For instance, the EU scaled system, 200 times smaller than the actual system, required more than 60Gb of RAM and 72-hour simulation requires a computation time of 10h and 15 minutes, performed on 3.6MHz-12CPUs (2 threads per core, 6 cores per socket) machine loaded at an average of 200%. A one year simulation would require almost the same amount of RAM (only information influencing the time step at stake is transferred), but would end after 52 days. Any trial to simulate the system at its real scale would require 200 times the 60Gb of RAM.

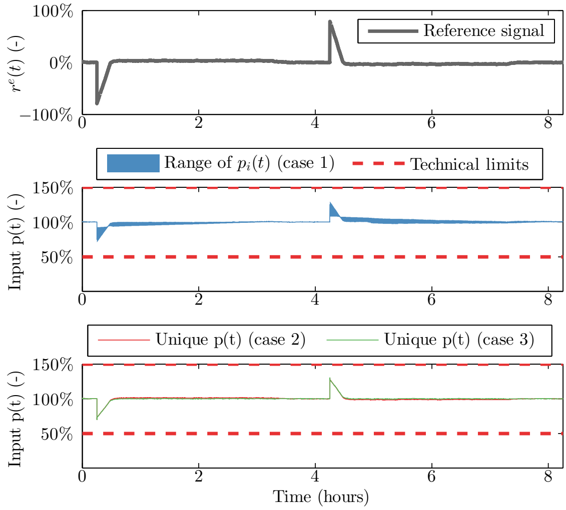


Figure 7.36: Evolution of the input computed autonomously by each load. Top chart: reference signal $\Delta f(t)/\phi$. Middle chart: input range $p_i(t)$ in case 1. Bottom chart: Unique input (no measurement error) in case 2 and 3. The input dispersion is due to the different run time impact in the inverse model and get limited as a consequence of the capacity allocation process.

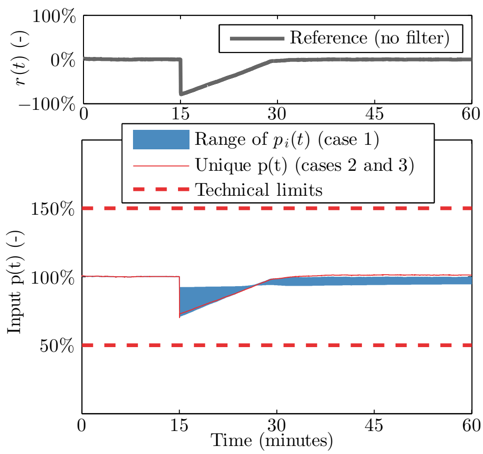


Figure 7.37: Evolution of the inputs computed autonomously by each load. Input dispersion originates from random run times and capacity allocation.

7.8.1 Aggregate models for all control cases

In order to make long term simulation accessible to any standard PC, the use of aggregate model can be useful. Let's define by $x_1(t), x_2(t), x_3(t)$ the expected value of the group response in the three control cases. Let Δp be the power deviation that any load can impose to its power rate $p(t) \in [1-\Delta p, 1+\Delta p]$. Then, the capacity allocation results in imposing a limit to the proportional gain $K_D^{dp}(\Delta p, T, a)$ that depends on this allowed power deviation, on the run time of the involved load T as well as on the filter parameter a. Let's recall the probability distribution Ω_T defined for $T \in [T_m, T_M]$ relative to the run time T at start, the average run time T_n at start, the average power of loads P_n. The average response droop offered by the group subject to the advanced control scheme is found below.

$$K_D^{av} = \frac{1}{T_n} \int_{T_m}^{T_M} K_D^{dp}(\Delta p, T) T \Omega_T \, dT \tag{7.95}$$

Then, the relative aggregate demand change of the load group can be computed.

Aggregate models in the three control cases

These aggregate models represent the controlled demand the group providing FCR with droop K_D in each considered case.

$$x_1(t) = \frac{K_D^{av}}{\phi}\left(\Delta f(t) - \frac{1}{a}\int_{t-a}^{t} \Delta f(s)\, ds\right) \tag{7.96}$$

$$\begin{aligned} x_2(t) = \frac{\Delta p}{\phi}\Bigg(&\Delta f(t) - \frac{1}{a}\int_{t-a}^{t} \Delta f(s)\, ds \\ &- \frac{1}{T_n}\int_{T_m}^{T_M} \Omega_T \int_{t-T}^{t} \left(\Delta f(s) - \frac{1}{a}\int_{s-a}^{s} \Delta f(z)\, dz\right) ds\Bigg) \end{aligned} \tag{7.97}$$

$$x_3(t) = \frac{\Delta p}{\phi}\left(\Delta f(t) - \frac{1}{T_n}\int_{T_m}^{T_M} \Omega_T \int_{t-T}^{t} \Delta f(s)\, ds\right) \tag{7.98}$$

The actual demand change $x_i(t) D_{0,i}$ depends on the number of involved loads. The provided FCR capacity FCR_i^{cap} in each control case is derived below.

$$FCR_1^{cap} = K_D^{av} D_{0,1} \quad FCR_2^{cap} = \Delta p D_{0,2} \quad FCR_3^{cap} = \Delta p D_{0,3} \tag{7.99}$$

Validation of the models

We firstly fix the number of loads, as was done in the above simulations. This is presented on figure 7.38. This figure can be compared to the individual simulation 7.27. Results are qualitatively speaking, very close in both cases.

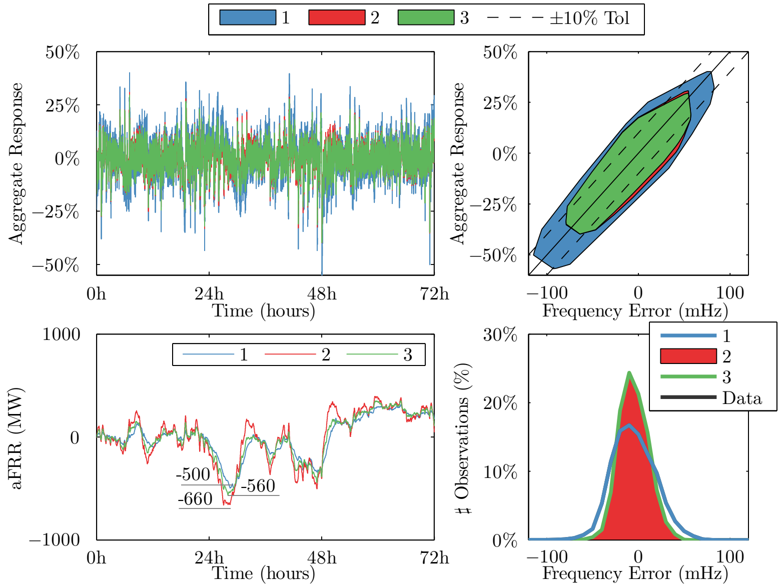


Figure 7.38: Simulation results with aggregate models. Groups have the same size which is scaled for the advanced control to provide the full FCR needs.

The impact parameter randomness and random start rates as well as the one of measurement errors cannot easily be estimated from an aggregate perspective. Indeed, the feedback effect that is inherent to FCR diminishes the impact of such errors on the frequency quality. Mainly aFRR volumes are impacted.

Scaling of the number of loads to have the same FCR

We may want to adapt $D_{0,i}$ for that in each case the full FCR requirements of the modeled system are covered by demand-side flexibility. This leads to the results on figure 7.39. As can be observed, the technical performances becomes very similar in all control cases, except for the aFRR requirements.

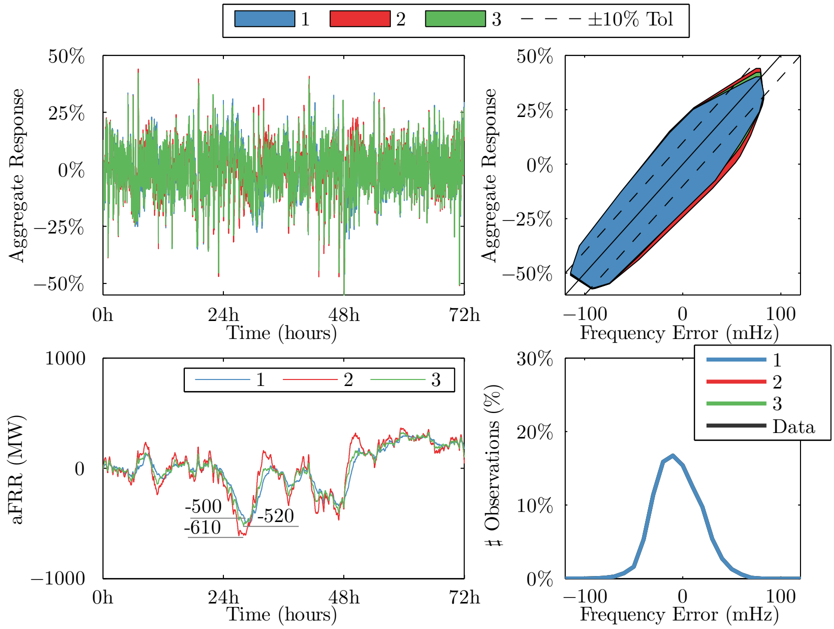


Figure 7.39: Simulation results with aggregate models. Groups have different size that are scaled to provide the same FCR.

From these results, the advanced scheme seems to be the best performer, but one must recall that the number of involved loads is increased by the capacity allocation process. Cost-based analysis is conducted below to take this element into consideration.

7.8.2 One year simulation

We conduct a one year simulation in which the aggregate models are exploited. The simulation lasted for less than 3 hours and required less than 0.9Gb of RAM what shows the clear advantage of exploiting aggregate models. The results are displayed on figures 7.40 (group performance) and 7.41 (frequency quality).

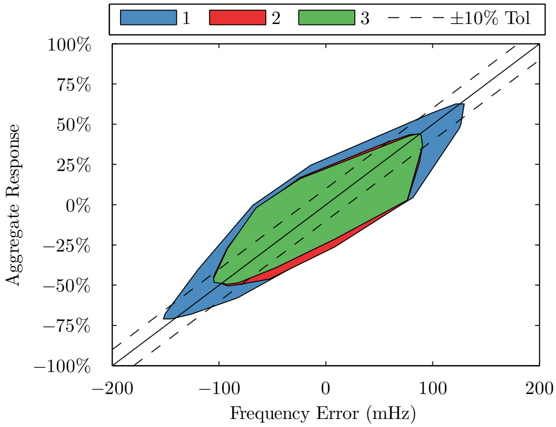


Figure 7.40: Perfomance of the group (aggregate models), 1y. simulation.

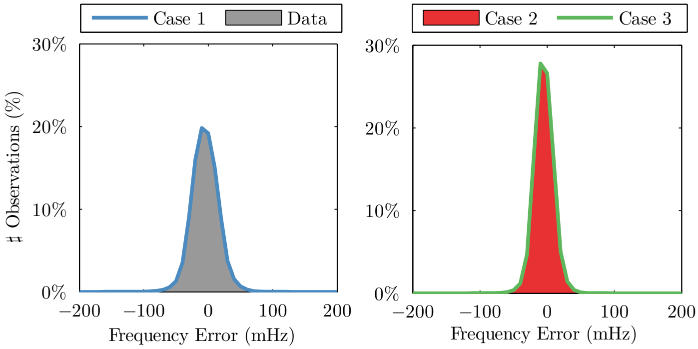


Figure 7.41: Frequency histograms, one year simulation (aggregate models).

Table 7.1, summarizes the upward ($aFRR^{up}$) and downward ($aFRR^{dn}$) needs for additional FRR obtained in the one year simulation. It also shows the amount of FCR_D actually provided by the group (fixed size) in each case. Comparing the last line of the table for case 1 (advanced) and 3 (simple) shows that the 36% loss of FCR capacity in the advanced framework leads to reduced needs for aFRR of about the same proportions.

Table 7.1: One year simulation : provided FCR_D and additional $aFRR$ needs.

	Case 1	Case 2	Case 3
FCR_D	3000 MW	4080 MW	4080 MW
	100%	136%	136%
$aFRR^{up}$	460 MW	730 MW	590 MW
	100%	160%	130%
$aFRR^{dn}$	-540 MW	-990 MW	-740 MW
	100%	180%	137%

7.8.3 Cost-benefit analysis

We conduct a similar cost-benefit analysis than performed in the simple control scheme (Fig. 6.28). Advanced controller costs limits are approximately $2/3^{rd}$ than the results obtained in the simple case.

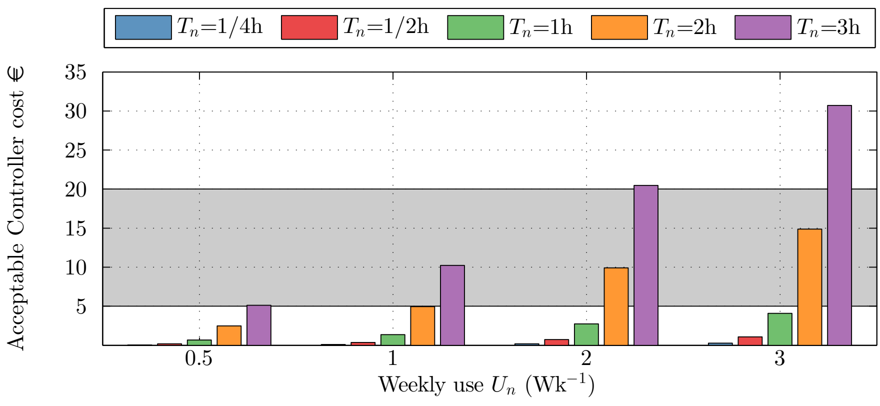
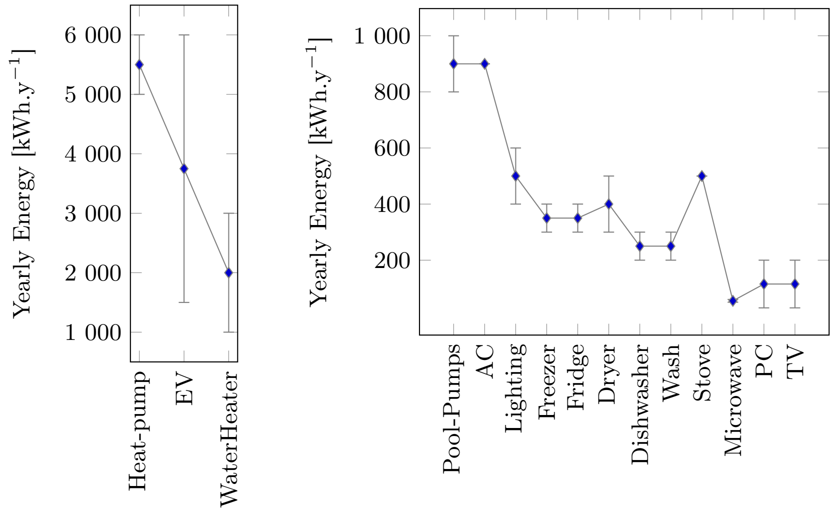


Figure 7.42: Maximum acceptable cost of the advanced power controller in case just the required amount of loads are equipped for replacing full FCR needs, as a function of the load's run time and weekly use. Parameters: $P_n = 2kW, K_D = 50\%$.

The results start being positive for loads consuming on average 200-300kWh per year and running during one consecutive hour. For comparison, the average annual energy consumption of common appliances in Belgian households is presented on figure 7.43.



(a) Larger appliances.

(b) Smaller appliances.

Figure 7.43: Annual energy consumption of residential appliances (Belgium).

Eligible residential appliances (advanced controller) are water heaters (night storage) and electric vehicles. The Wet appliances (Dryer, Washing Machine, Dishwashers) are of interest but their controller should be adapted to take their cycle profile into account (see e.g., [144]).

Chapter 8

Conclusions

Energy Constrained Loads (ECLs) can contribute to Frequency Containment Reserves (FCR) in an efficient way from an overall system perspective both from a technical and an economic point of view. The main driver of the control profitability is to restrict implementation to the loads that are the most interesting, and only to them. This is what limits the overall cost expenditure and leads to the best impact on the overall system. In today's context, flexibility is not yet rare enough to justify the control of the system's tiniest loads.

On the end user's side, the picture is probably less attractive. End users cannot expect to earn enough to reduce their energy bill significantly. However, the control impact on their daily use of energy is also very limited. There are therefore two ways ahead. For the largest loads (e.g., electric vehicles), one could afford to let each consumer choose whether to accept external control. In short, these appliances generate enough profits. Some of it can be dedicated to the process of convincing the end-user (e.g., marketing costs). For the smaller ones (e.g., batch water heating in different appliances), only mandatory and massive implementation programs and standardized controllers can be afforded.

As implementation will take time, the most interested parties, namely system operators and policy makers, should probably think about it early enough. A few elements could prevent the deployment of simple power controllers for FCR participation. Firstly, long-term historical simulations have shown that the power system operations would be impacted by a massive use of ECL as unique provider of FCR. Mostly, the operator will need to keep track of the frequency integral and take actions preemptively to efficiently managing its impacts on the load control performances and profitability. Secondly, technical requirements on FCR may be overly restrictive and will need to be adapted.

8.1 Main findings

We summarize here below the key elements discussed in this work. We refer to the exact chapter in which results are discussed in braces.

- Historically, small electric loads have been involved in some specific system operations each time their use consisted in a solution more valuable than reachable alternatives. In general, loads are efficient to provide strong response in case of emergency situations, or in rare occasions. This occurs in systems whose demand takes large shares of weather-sensitive loads (e.g., air-conditioning), systems with peculiar topology leading to large instability issues, systems in which the generation mix experiences seasonal variability (e.g., hydro-power) leading to adequacy problems, and systems that rely on a few large infrastructures whose failures threaten its security (e.g., connection tie-line for import). [Chap. 2]

- Small loads do not participate, yet, in Frequency Containment Reserves, except in a few pilot projects. Mostly, this can be explained by the fact that their contribution starts being effective when implemented on a large scale. The long-term success of such programs requires a strong political willingness. In addition, relatively complex controllers (i.e., frequency measurement) must be available at low-cost. [Chap. 3]

- In the literature, studies show a general trend of neglecting the overall system perspective of introducing ECL within system operations. This is especially true in the case of long-term consequences (frequency quality, etc.). [Chap. 3]

- The consumption of a large group of small loads experiences some variability which is a direct consequence of the uncoordinated behavior of the involved loads (random starting time) and of the dispersion of their technical parameters. For a group counting N *running* loads, the relative variations of demand around its expected value stays in general below $3/\sqrt{N}$. For this variability to be negligible in regards to the tolerance applied on control performances, the number of running loads should be about 300 to 500. As the considered loads run for short period of time, the total amount of involved loads must be much higher. Considering average appliance use, about 100.000 individual loads should get involved in the control scheme. This calls for massive implementation programs. [Chap. 4]

- The demand of large groups of ECL cannot be controlled at will. The consequence of the load's energy constraint is that the group is only able to shift energy across time. The energy that is not consumed at some point in time will be recovered later on, a phenomenon known as the energy rebound. This rebound can take many different shapes according to the way loads are controlled. Yet, one main parameter of influence is the load's initial run time, that is the time period during which the load was supposed to consume energy in case no control action had been undertaken. [Chap. 5]

- The ECL's power consumption can be controlled in an autonomous setting and respond with great precision to frequency changes. Loads need some prior knowledge about their own behavior, in particular their scheduled run time. In both short-run and long-run simulations, the simplest controllers, which have no means of counteracting the energy rebound, are exhibiting acceptable performances. [Chap. 6]. This is conditioned to the fact that slower flexibility resources can compensate for the negative impact of the energy rebound. The additional requirements in the amount of slower resources that must be specifically dedicated to rebound management strongly depends on the loads' run time. The capacity increase is moderate as soon as loads run for one consecutive hour. [Chap. 6]

- The simpler controllers performances are dependent on the fact that slower flexibility resources can be compensated for their rebound error. Without correctly functioning of the slower resources, they would not behave correctly. If the system operators consider their behavior as acceptable, they should adapt the current prequalification processes used to certify. But this may be overlooking important risk considerations? [Chap. 6]

- We proposed aggregate models which represent accurately the behavior of the group of controlled loads under two conditions: the load should be large and the energy content of the service it delivers should be limited. These models take into account the rebound error, and boil down to apply a form of high-pass filtering on the reference signal they are supposed to track. They have a simple mathematical form (linear) and are particularly useful to conduct long-term simulations. [Chap. 5]

- Economic analyses have shown that the generated profits would be in the 0.1-20€ range for most household appliances. In fact, the generated profits are, under our assumptions, about 0.8¢€/kWh, of which controller costs must be subtracted. In Belgium, this corresponds to 4% of the total electricity price for residential consumers. Such profit level seems acceptable as soon as loads consume 200-300kWh/y and run for at least 30 consecutive minutes. Indeed, the main driver of profitability is to limit the number of loads, and therefore to target only those able to contribute more annually, that is the ones with sufficient energy consumption. [Chap. 6]

- ECLs can also be controlled in a more advanced setting in which the group compensates for its own energy rebound. They become therefore able to guarantee a certain level of service. The consequence is that part of the initially available power flexibility must be kept for this rebound management and cannot be sold to the system operator. At the same time, they decrease the reliance on slower resources which is beneficial to the system as a whole. Yet, the conservative approach proposed in this work leads to a decreased profitability. [Chap. 7]

- The system operator's role in increasing the control profitability is significant. Indeed, the type of slower resources exploited for rebound management can vary. In the most restrictive case, the participating resources are the automatically activated frequency restoration reserves (aFRR) which are the most expensive resources at the disposal of the system operator. The aFRR are activated in a *reactive* way, meaning that they have to wait for something to occur before they start compensating for it. The system operator (SO) however has much cheaper resources at hand for helping aFRR performing their task within the balancing market. The involved resources are in general much slower which means that the system operator must forecast the short-term evolution of its system in order to make an efficient use of those resources. As the rebound error can partially be predicted from the frequency integral, preventive actions will become an important element of daily system operations. [Chap. 6 and 7]

8.2 Next research paths

Here are some suggestions for further exploration and improved integration of ECLs within system operations.

- Explore the exact EU potential. What are the exact ECL parameters that would lead to year-long and day-long availability (cross-sector) ? The proposed aggregate models are suited to exploit time-varying parameter distributions.

- Detailed analysis on the end-user consequences of the ECL control. In particular, the controlled policy consisting in rapid On and Off switching deserves attention. The controller should try to avoid rapid switching or oscillatory behavior. Furthermore, as the control is selective, all loads do not react in the same way. This induces a certain dispersion of the end-user consequences, that should be explored in more details. Several studies have looked at similar issues in the case of TCL, and their results could probably be adapted to the ECL case.

- The advanced controllers should be adapted to lead to less conservative approaches. The technical and economic aspects can be integrated into the controller design process. In particular, loads should maximize the immediate responsiveness, as it has a great influence in limiting extreme frequency excursions. Consequently, the integral of the frequency excursions is limited, and so is the rebound error. In the process, the coordination with other system operations is crucial. As loads will still need to recover their initial state, they must progressively, as fast as possible, reduce their response magnitude. If this is done in coordination with cheap preventive actions undertaken by the system operator, the frequency integral will stay unchanged, and the overall profitability will be increased.

- Optimize the energy recovery process. Instead of relying on expensive aFRR, the preemptive actions of the SO should be optimized to exploit the flexibility of the balancing market. Stochastic economic dispatch is probably a very important tool to conduct such an optimization.

- One of the most important: some practical demonstration is needed.

Bibliography

[1] "A talk by Nikola Tesla". In: *The New York Times* (May 1891). URL: `http://www.teslauniverse.com`.

[2] M. Alizadeh, G. Kesidis, and A. Scaglione. "Clustering consumption in queues: A scalable model for electric vehicle scheduling". In: *Signals, Systems and Computers, 2013 Asilomar Conference on.* Nov. 2013, pp. 374–378. DOI: `10.1109/ACSSC.2013.6810299`.

[3] M. Alizadeh, A. Scaglione, A. Applebaum, G. Kesidis, and K. Levitt. "Reduced-Order Load Models for Large Populations of Flexible Appliances". In: *Power Systems, IEEE Transactions on* 30.4 (July 2015), pp. 1758–1774. ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2354345`.

[4] Fernando L Alvarado, Jianping Meng, Christopher L DeMarco, and Wellington S Mota. "Stability analysis of interconnected power systems coupled with market dynamics". In: *Power Systems, IEEE Transactions on* 16.4 (2001), pp. 695–701.

[5] Monitoring Analytics. *PJM - State of the Market Report.* Dec. 2015. URL: `http://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2015.shtml`.

[6] Göran Andersson. *Dynamics and control of electric power systems.* ETH Zürich, 2012. URL: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.7315&rep=rep1&type=pdf`.

[7] S.-L. Andersson, A.K. Elofsson, M.D. Galus, L. Goransson, S. Karlsson, F. Johnsson, and G. Andersson. "Plug-in hybrid electric vehicles as regulating power providers: Case studies of Sweden and Germany". In: *Energy Policy* 38.6 (2010), pp. 2751–2762. DOI: `http://dx.doi.org/10.1016/j.enpol.2010.01.006`.

[8] M Andreolas. "Mega load management system pays dividends". In: *Transmission & Distribution World* (2004), pp. 33–37.

[9] D. Angeli and P.-A. Kountouriotis. "A Stochastic Approach to Dynamic Demand Refrigerator Control". In: *Control Systems Technology, IEEE Transactions on* 20.3 (May 2012), pp. 581–592. ISSN: 1063-6536. DOI: `10.1109/TCST.2011.2141994`.

[10] United States Court of Appeals for the district of Columbia Circuit. "On Petitions for Review of Orders of the Federal Energy Regulatory Commission". In: *FERC Stats. & Regs. No. 11-1486.* 2014, May 23.

[11] A. Arteconi, N.J. Hewitt, and F. Polonara. "State of the art of thermal storage for demand-side management". In: *Applied Energy* 93 (2012). (1) Green Energy; (2)Special Section from papers presented at the 2nd International Enery 2030 Conf, pp. 371–389. ISSN: 0306-2619. DOI: `http://dx.doi.org/10.1016/j.apenergy.2011.12.045`. URL: `http://www.sciencedirect.com/science/article/pii/S0306261911008415`.

[12] A. D. Barbour and O. Chryssaphinou. "Compound Poisson Approximation: A User's Guide". In: *The Annals of Applied Probability* 11.3 (2001), pp. 964–1002. ISSN: 10505164. URL: `http://www.jstor.org/stable/2699886`.

[13] S. Bashash and H.K. Fathy. "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads". In: *American Control Conference (ACC), 2011*. June 2011, pp. 4546–4553. DOI: `10.1109/ACC.2011.5990939`.

[14] William J Baumol and David F Bradford. "Optimal departures from marginal cost pricing". In: *The American Economic Review* 60.3 (1970), pp. 265–283.

[15] Charles Crisp Benton. "Off-peak cooling using phase change material". Master Thesis. Massachusetts Institute of Technology, 1979. URL: `http://hdl.handle.net/1721.1/71312`.

[16] R.J. Bessa and M.A. Matos. "Optimization models for an {EV} aggregator selling secondary reserve in the electricity market". In: *Electric Power Systems Research* 106 (2014), pp. 36–50. ISSN: 0378-7796. DOI: `http://dx.doi.org/10.1016/j.epsr.2013.08.006`. URL: `http://www.sciencedirect.com/science/article/pii/S0378779613002162`.

[17] B. Biegel, L.H. Hansen, P. Andersen, and J. Stoustrup. "Primary Control by ON/OFF Demand-Side Devices". In: *Smart Grid, IEEE Transactions on* 4.4 (Dec. 2013), pp. 2061–2071. ISSN: 1949-3053. DOI: `10.1109/TSG.2013.2257892`.

[18] Katie Bloor, Andrew Howe, Alessandra Suardi, and Stefano Frattesi. *Dynamic Ancillary Service Provided By Loads With Inherent Energy Storage*. Tech. rep. RLtec, 2009.

[19] Josh L. Bode, Michael J. Sullivan, Dries Berghman, and Joseph H. Eto. "Incorporating residential {AC} load control into ancillary service markets: Measurement and settlement". In: *Energy Policy* 56 (2013), pp. 175–185. DOI: `http://dx.doi.org/10.1016/j.enpol.2012.12.024`.

[20] Roger E Bohn, Michael C Caramanis, and Fred C Schweppe. *Optimal Spot Pricing of Electricity: Theory*. Massachusetts Institute of Technology, Energy Laboratory, 1981.

[21] Severin Borenstein and James Bushnell. *The U.S. Electricity Industry After 20 Years of Restructuring*. Tech. rep. National Bureau of Economic Research, Apr. 2015. URL: https://ei.haas.berkeley.edu/research/papers/WP252.pdf.

[22] T. Borsche and G. Andersson. “A review of demand response business cases”. In: *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*. Oct. 2014, pp. 1–6.

[23] T. Borsche, A. Ulbig, M. Koller, and G. Andersson. “Power and energy capacity requirements of storages providing frequency control reserves”. In: *Power and Energy Society General Meeting (PES), 2013 IEEE*. July 2013, pp. 1–5. DOI: 10.1109/PESMG.2013.6672843.

[24] Theodor Borsche, Uros Markovic, and Goran Andersson. “A new algorithm for primary frequency control with cooling appliances”. English. In: *Computer Science, Research and Development* (2014), pp. 1–7. DOI: 10.1007/s00450-014-0289-1. URL: http://dx.doi.org/10.1007/s00450-014-0289-1.

[25] Martin Braun and Philipp Strauss. “A review on aggregation approaches of controllable distributed energy units in electrical power systems”. In: *International Journal of Distributed Energy Resources* 4.4 (2008), pp. 297–319.

[26] Richard E. Brown and Jonathan G. Koomey. *Electricity Use in California Past Trends and Present Usage Patterns*. 2002. URL: http://enduse.lbl.gov/Info/LBNL-47992.pdf.

[27] James Bushnell, Benjamin F. Hobbs, and Frank A. Wolak. “When It Comes to Demand Response, Is {FERC} Its Own Worst Enemy?” In: *The Electricity Journal* 22.8 (2009), pp. 9–18.

[28] D.S. Callaway and IA Hiskens. “Achieving Controllability of Electric Loads”. In: *Proceedings of the IEEE* 99.1 (Jan. 2011), pp. 184–199. DOI: 10.1109/JPROC.2010.2081652.

[29] Peter Cappers, Charles Goldman, and David Kathan. “Demand response in U.S. electricity markets: Empirical evidence”. In: *Energy* 35.4 (2010), pp. 1526–1535.

[30] M. Caramanis. “Investment Decisions and Long-Term Planning Under Electricity Spot Pricing”. In: *Power Apparatus and Systems, IEEE Transactions on* PAS-101.12 (Dec. 1982), pp. 4640–4648. ISSN: 0018-9510. DOI: 10.1109/TPAS.1982.317321.

[31] Fidel Castro Rodriguez. *Wright tariffs in the spanish electricity industry: The case of residential consumption.* Tech. rep. Universidad Carlos III de Madrid, 1997.

[32] D.P. Chassin, P. Du, and J.C. Fuller. "The potential and limits of residential demand response control strategies". In: *Power and Energy Society General Meeting, 2011 IEEE.* July 2011, pp. 1–6. DOI: 10.1109/PES.2011.6039719.

[33] Climact and Vito. *Scenarios for a low carbon belgium by 2050.* Tech. rep. SPF environnement, 2013.

[34] Federal Energy Regulatory Commission. "Demand Response Compensation in Organized Wholesale Energy Markets, Order No. 745." In: *FERC Stats. & Regs. 31,322.* 2011.

[35] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth. "On the Lambert W function". English. In: *Advances in Computational Mathematics* 5.1 (1996), pp. 329–359. ISSN: 1019-7168. DOI: 10.1007/BF02124750.

[36] Peter Cramton and Steven Stoft. *The Convergence of Market Designs for Adequate Generating Capacity.* White Paper. University of Maryland, Department of Economics - Peter Cramton, 2006. URL: http://EconPapers.repec.org/RePEc:pcc:pccumd:06mdfra.

[37] A Dames, K Roussopoulos, T Norris, and A Howe. "Smart Responsive Electrical Load". US 8912683 B2. Open Energi Limited. Dec. 16, 2014.

[38] Cedric De Jonghe. "Short-term demand response in elecitricy generation planning and scheduling : facilitating wind power integration". PhD thesis. Katholieke Universiteit Leuven Faculteit Ingenieurswetenschappen, 2011.

[39] Todd De Wayne. "Alcoa - Demand Response Innovation". In: *FERC Technical Conference on Frequency Regulation Compensation.* 2010.

[40] S. Deml, A. Ulbig, T. Borsche, and G. Andersson. "The role of aggregation in power system simulation". In: *PowerTech, 2015 IEEE Eindhoven.* June 2015, pp. 1–6. DOI: 10.1109/PTC.2015.7232755.

[41] DENA - Deutsche Energie-Agentur. *Dena Grid Study II - Integration of Renewable Energy Sources in the German Power Supply System from 2015-2020 with an Outlook to 2025.* Tech. rep. DENA - Deutsche Energie-Agentur, 2010.

[42] Dep. of commerce. *State of Washington. Washington State Energy Assurance and Emergency Preparedness Plan.* 2013. URL: http://www.commerce.wa.gov/Documents/Energy-Emergency-Plan-2013.pdf.

[43] J. Donadee and M. Ilic. "Stochastic co-optimization of charging and frequency regulation by electric vehicles". In: *North American Power Symposium (NAPS), 2012*. Sept. 2012, pp. 1–6. DOI: 10.1109/NAPS.2012.6336373.

[44] P.J. Douglass, R. Garcia-Valle, P. Nyeng, J. Ostergaard, and M. Togeby. "Smart Demand for Frequency Regulation: Experimental Results". In: *Smart Grid, IEEE Transactions on* 4.3 (Sept. 2013), pp. 1713–1720. ISSN: 1949-3053. DOI: 10.1109/TSG.2013.2259510.

[45] B. Ebrahimi and J. Mohammadpour. "Aggregate modeling and control of plug-in electric vehicles for renewable power tracking". In: *American Control Conference (ACC), 2014*. June 2014, pp. 2035–2040. DOI: 10.1109/ACC.2014.6859169.

[46] EDF. *Tarif tempo pour particuliers*. last visited in october 2015. URL: http://particuliers.edf.com.

[47] Elia System Operator. *Ancillary Services: Volumes & Prices*. 2015. URL: http://www.elia.be/en/suppliers/purchasing-categories/energy-purchases/Ancillary-Services-Volumes-Prices.

[48] Elia System Operator, Febeliec, and EnergyVille. *Demand Response Survey - Summary results*. 2013.

[49] EnerNOC. *Alberta*. Tech. rep. EnerNoc, 2012. URL: http://www.enernoc.com.

[50] W.A. England and L.W. Harrison. "Integrating the Load Management Function into the Energy Control Center A Case History". In: *Power Apparatus and Systems, IEEE Transactions on* PAS-104.6 (June 1985), pp. 1281–1285.

[51] ENTSO-E. *ENTSO-E Network Code on Demand Connection*. 2012. URL: http://www.entsoe.eu.

[52] ENTSO-E. *Network Code on Load-Frequency control and reserve*. 2013. URL: http://www.entsoe.eu.

[53] ENTSO-E. *System adequacy forecast 2011-2025*. 2013.

[54] Antonio Estache and Anne-Sophie Steichen. "Is Belgium overshooting in its policy support to cut the cost of capital of renewable sources of energy ?" In: *Reflets et perspectives de la vie économique* Tome LIV.2015/1-2 (2015). Ed. by De Boeck Supérieur, p. 236. DOI: 10.3917/rpve.541.0033.

[55] Joseph Eto. *The past, present, and future of US utility demand-side management programs*. Tech. rep. Lawrence Berkeley National Lab., CA (United States)., 1996.

[56] European Parliament and Council of the European Union. *Directive 2009/72/EC concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC*. Brussels, July 2009. URL: `http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0072&rid=1`.

[57] Conrad Eustis and Gale Horst. "Appliance Interface for Grid Responses". In: *Proc. of the Grid Interop Forum*. 2007, pp. 7–9.

[58] Ahmad Faruqui and Lisa Wood. *Quantifying the benefits of dynamic pricing in the mass market*. Tech. rep. Edison Electric Institute, The Brattle Group, 2008.

[59] T. Fawcett, K. Lane, B. Boardman, and *et. al. Lower Carbon futures for european households*. 2000. URL: `http://www.eci.ox.ac.uk/research/energy/downloads/lowercarbonfuturereport.pdf`.

[60] Frank A. Wolak. "Diagnosing the California Electricity Crisis". In: *The Electricity Journal* (2003). Ed. by Elsevier. DOI: `10.1016/S1040-6190(03)00099-X`.

[61] F. Friend. "Cold load pickup issues". In: *Protective Relay Engineers, 2009 62nd Annual Conference for*. Mar. 2009, pp. 176–187. DOI: `10.1109/CPRE.2009.4982512`.

[62] Matthias D. Galus, S. Koch, and G. Andersson. "Provision of Load Frequency Control by PHEVs, Controllable Loads, and a Cogeneration Unit". In: *Industrial Electronics, IEEE Transactions on* 58.10 (Oct. 2011), pp. 4568–4582. DOI: `10.1109/TIE.2011.2107715`.

[63] Hans Christian Gils. "Assessment of the theoretical demand response potential in Europe". In: *Energy* 67 (2014), pp. 1–18. ISSN: 0360-5442. DOI: `http://dx.doi.org/10.1016/j.energy.2014.02.019`. URL: `http://www.sciencedirect.com/science/article/pii/S0360544214001534`.

[64] M. Gonzalez Vaya and G. Andersson. "Combined smart-charging and frequency regulation for fleets of plug-in electric vehicles". In: *Power and Energy Society General Meeting (PES), 2013 IEEE*. July 2013, pp. 1–5. DOI: `10.1109/PESMG.2013.6672852`.

[65] C.M. Grinstead and J.L. Snell. *Introduction to Probability*. American Mathematical Society, 1997. ISBN: 9780821807491.

[66] Donald J Hammerstrom, Jerry Brous, David P Chassin, Gale R Horst, Robert Kajfasz, Preston Michie, Terry V Oliver, Teresa A Carlon, Conrad Eustis, Olof M Jarvegren, and Miscs. *Pacific Northwest GridWise Testbed Demonstration Projects; Part II. Grid Friendly Appliance Project*. Tech. rep. Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2007.

[67] Hancher et al. *Shift, Not Drift. Towards Active Demand Response and Beyond - Final Report.* THINK FP7 project, 2013. URL: http://www.eui.eu/Projects/THINK/Research/Topic11.aspx.

[68] B.F. Hastings. "Ten Years of Operating Experience with a Remote Controlled Water Heater Load Management System at Detroit Edison". In: *Power Apparatus and Systems, IEEE Transactions on* PAS-99.4 (July 1980), pp. 1437–1441. ISSN: 0018-9510. DOI: 10.1109/TPAS.1980.319566.

[69] G. Heffner, C. Goldman, B. Kirby, and M. Kintner-Meyer. *Loads providing ancillary services: Review of international experience.* Tech. rep. Lawrence Berkeley National Laboratory, 2007.

[70] K. Heussen, S. You, B. Biegel, L.H. Hansen, and K.B. Andersen. "Indirect control for demand side management - A conceptual introduction". In: *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on.* Oct. 2012, pp. 1–8. DOI: 10.1109/ISGTEurope.2012.6465858.

[71] Eric Hirst, Ralph Cavanagh, and Peter Miller. "The future of DSM in a restructured US electricity industry". In: *Energy Policy* 24.4 (1996), pp. 303–315.

[72] M.D. Ilic, N. Popli, Jhi-Young Joo, and Yunhe Hou. "A possible engineering and economic framework for implementing demand side participation in frequency regulation at value". In: *Power and Energy Society General Meeting, 2011 IEEE.* 2011, pp. 1–7. DOI: 10.1109/PES.2011.6039498.

[73] S. Izadkhast, P. Garcia-Gonzalez, and P. Frias. "An Aggregate Model of Plug-In Electric Vehicles for Primary Frequency Control". In: *Power Systems, IEEE Transactions on* 30.3 (May 2015), pp. 1475–1482. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2337373.

[74] Paul L Joskow. "Contributions to the theory of marginal cost pricing". In: *The Bell Journal of Economics* 7.1 (1976), pp. 197–206.

[75] Paul L. Joskow and Donald B. Marron. "What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs". In: *The Energy Journal* Number 4 (1992), pp. 41–74. URL: http://ideas.repec.org/a/aen/journl/1992v13-04-a03.html.

[76] Willett Kempton, Victor Udo, Ken Huber, Kevin Komara, Steve Letendre, Scott Baker, Doug Brunner, and Nat Pearre. *A test of vehicle-to-grid (V2G) for energy storage and frequency regulation in the PJM system. Results from an Industry-University Research Partnership.* Tech. rep. University of Delaware, 2008. URL: http://www.udel.edu/V2G/resources/test-v2g-in-pjm-jan09.pdf.

[77] Jin-Ho Kim and Anastasia Shcherbakova. “Common failures of demand response”. In: *Energy* 36.2 (2011), pp. 873–880. ISSN: 0360-5442. DOI: http://dx.doi.org/10.1016/j.energy.2010.12.027. URL: http://www.sciencedirect.com/science/article/pii/S0360544210007176.

[78] M. Klobasa. “Analysis of demand response and wind integration in Germany’s electricity market”. English. In: *IET Renewable Power Generation* 4 (1 Jan. 2010), 55–63(8). URL: http://digital-library.theiet.org/content/journals/10.1049/iet-rpg.2008.0086.

[79] M. Klobasa. “Dynamische Simulation eines Lastmanagements und Integration von Windenergie in ein Elektrizitätsnetz auf Landesebene unter regelungstechnischen und Kostengesichtspunkten”. PhD thesis. Eidgenössische Technische Hochschule Zürich, 2007.

[80] M. Klobasa, C. Obersteiner, M. Ragwitz, and H. AVER. *Stategies for an efficient integration of wind power considering demand response.* Tech. rep. Energy Economics Group-Vienna University of Technology, 2007.

[81] Stephan Koch, Johanna L Mathieu, and Duncan S Callaway. “Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services”. In: *Proceedings of the 17th Power Systems Computation Conference (PSCC).* Aug. 22–26, 2011, pp. 1–7.

[82] J. K. Kok, C. J. Warmer, and I. G. Kamphuis. “PowerMatcher: Multiagent Control in the Electricity Infrastructure”. In: *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems.* AAMAS 05. The Netherlands: ACM, 2005, pp. 75–82. DOI: 10.1145/1082473.1082807.

[83] V Kolmanovskii and A Myshkis. *Applied theory of functional differential equations.* Springer, 1992.

[84] J. Kondoh. “Direct load control for wind power integration”. In: *Power and Energy Society General Meeting, 2011 IEEE.* IEEE. 2011, pp. 1–8.

[85] J. Kondoh, N. Lu, and D.J. Hammerstrom. “An evaluation of the water heater load potential for providing regulation service”. In: *Power and Energy Society General Meeting, 2011 IEEE.* IEEE. 2011, pp. 1–8.

[86] H.J. Kunisch, K.G. Kramer, and H. Dominik. "Battery Energy Storage Another Option for Load-Frequency-Control and Instantaneous Reserve". In: *Energy Conversion, IEEE Transactions on* EC-1.3 (Sept. 1986), pp. 41–46. ISSN: 0885-8969. DOI: 10.1109/TEC.1986.4765732.

[87] Venkatachalam Lakshmanan, Mattia Marinelli, Anna M. Kosek, Per B. Nørgård, and Henrik W. Bindner. "Impact of thermostatically controlled loads' demand response activation on aggregated power: A field experiment". In: *Energy* 94 (2016), pp. 705–714. ISSN: 0360-5442. DOI: http://dx.doi.org/10.1016/j.energy.2015.11.050. URL: http://www.sciencedirect.com/science/article/pii/S036054421501600X.

[88] A. Latiers, E. De Jaeger, and F. Glineur. "Toward decentralized primary control from demand response". In: *CIGRE 2014 belgium symposium*. 2014.

[89] Arnaud Latiers. "Demand Response Perspectives for Belgium". In: *Reflets et perspectives de la vie économique* LIV.1-2 (2015), pp. 185–203. ISSN: 0034-2971. DOI: 10.3917/rpve.541.0185.

[90] Arnaud Latiers, François Glineur, and Emmanuel De Jaeger. "Residential demand management and distribution grid impact assessment." In: *CIRED Lyon Conference*. 2015.

[91] Arnaud Latiers and Cédric Saussez. "Smart-Grids: Modèle et gestion d'un quartier résidentiel". Université Catholique de Louvain, 2010.

[92] Andreas Latzerath. "High Frequency Power Modulation - TRIMET Smelters Provide Primary Control Power for Stabilizing the Frequency in the Electricity Grid". In: *Light Metals 2013*. John Wiley & Sons, Inc., 2013, pp. 659–662. DOI: 10.1002/9781118663189.ch112. URL: http://dx.doi.org/10.1002/9781118663189.ch112.

[93] Micheal Lee. *Assessment of Demand Response and Advanced Metering (Staff Report)*. 2014. URL: http://www.ferc.gov/legal/staff-reports/2014/demand-response.pdf.

[94] Larry Leemis. *Estimating and Simulating Nonhomogeneous Poisson Processes*. Department of Mathematics, The College of William & Mary. 2003. URL: http://www.math.wm.edu/~leemis/icrsa03.pdf.

[95] S. Lefebvre and C. Desbiens. "Residential load modeling for predicting distribution transformer load behavior, feeder load and cold load pickup". In: *International Journal of Electrical Power & Energy Systems* 24.4 (2002), pp. 285–293. ISSN: 0142-0615. DOI: http://dx.doi.org/10.1016/S0142-0615(01)00040-0. URL: http://www.sciencedirect.com/science/article/pii/S0142061501000400.

[96] P. Lewis, C. Dromacque, S. Brennen, J. Stromback, and D. Kennedy. *EMPOWER DEMAND 2 - Energy Efficiency through Information and Communication Technology - Best Practice Examples and Guidance*. VasaaEtt, Esmig, 2012. URL: http://esmig.eu/sites/default/files/final_empower_2_demand_report_final_distr2.pdf.

[97] Linear Project. *Demand Response for families*. Project Report. 2014. URL: http://www.linear-smartgrid.be/sites/default/files/Linear%20Final%20Report%20-%20lr2.pdf.

[98] Hui Liu, Zechun Hu, Yonghua Song, and Jin Lin. "Decentralized Vehicle-to-Grid Control for Primary Frequency Regulation Considering Charging Demands". In: *Power Systems, IEEE Transactions on* 28.3 (Aug. 2013), pp. 3480–3489. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2013.2252029.

[99] N. Lu, M. R. Weimar, Y.V. Makarov, F. J. Rudolph, S. N. Murthy, J. Arseneaux, and C. Loutan. "Evaluation of the flywheel potential for providing regulation service in California". In: *Power and Energy Society General Meeting, 2010 IEEE*. July 2010, pp. 1–6. DOI: 10.1109/PES.2010.5590114.

[100] Marcel Maris, Andre Jaumotte, and Pierre Govaerts. *Histoire du nucleaire en Belgique, 1990-2005*. Peter Lang international scientific edition, 2008.

[101] V.J. Martinez and H. Rudnick. "Design of Demand Response programs in emerging countries". In: *Power System Technology (POWERCON), 2012 IEEE International Conference on*. Oct. 2012, pp. 1–6. DOI: 10.1109/PowerCon.2012.6401387.

[102] T. Masuta and A. Yokoyama. "Supplementary Load Frequency Control by Use of a Number of Both Electric Vehicles and Heat Pump Water Heaters". In: *Smart Grid, IEEE Transactions on* 3.3 (Sept. 2012), pp. 1253–1262. ISSN: 1949-3053. DOI: 10.1109/TSG.2012.2194746.

[103] J. Mathieu, M. Dyson, D. Callaway, and A. Rosenfeld. "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations." In: *Proceedings of the ACEEE Summer Study on Buildings*. 2012.

[104] JL Mathieu. "Modeling, Analysis, and Control of Demand Response Resources". PhD thesis. University of California, Berkeley, 2012.

[105] J.M. Mcintyre, L. Ciecior, A. Kaspar, and D. Castrop. "Distributed Intelligence in Load Control: Results of an Experiment Using Demand Limiting Devices for Residential Load Control". In: *Power Apparatus and Systems, IEEE Transactions on* PAS-104.5 (May 1985), pp. 1140–1146. ISSN: 0018-9510. DOI: 10.1109/TPAS.1985.323465.

[106] Aimee T. McKane, Ivin Rhyne, Alexander B. Lekov, Lisa Thompson, and Mary Ann Piette. "Automated Demand Response: The Missing Link in the Electricity Value Chain". In: *2008 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA, 2008.

[107] O. Megel, J.L. Mathieu, and G. Andersson. "Maximizing the potential of energy storage to provide fast frequency control". In: *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*. Oct. 2013, pp. 1–5.

[108] S. Meyn, P. Barooah, A. Busic, Y. Chen, and J. Ehren. "Ancillary Service to the Grid Using Intelligent Deferrable Loads". In: *Automatic Control, IEEE Transactions on* PP.99 (2015), pp. 1–1. ISSN: 0018-9286. DOI: 10.1109/TAC.2015.2414772.

[109] Wim Michiels and Silviu-Iulian Niculescu. *Stability, Control, and Computation for Time-Delay Systems: An Eigenvalue-Based Approach*. Vol. 27. SIAM, 2014.

[110] Noel M. Mirasol. "The Output of an M/G/? Queuing System is Poisson". English. In: *Operations Research* 11.2 (1963), pp. 282–284. ISSN: 0030364X. URL: http://www.jstor.org/stable/168236.

[111] A. Molina-García, F. Bouffard, and D.S. Kirschen. "Decentralized demand-side contribution to primary frequency control". In: *Power Systems, IEEE Transactions on* 26.1 (2011), pp. 411–419.

[112] Yunfei Mu, Jianzhong Wu, J. Ekanayake, N. Jenkins, and Hongjie Jia. "Primary Frequency Response From Electric Vehicles in the Great Britain Power System". In: *Smart Grid, IEEE Transactions on* 4.2 (June 2013), pp. 1142–1150. ISSN: 1949-3053. DOI: 10.1109/TSG.2012.2220867.

[113] Zoran Nenadic. *Poisson Process, Spike Train and All That*. Division of Engineering and Applied Science, California Institute of Technology. 2002. URL: http://robotics.caltech.edu/~zoran/Research/poisson/poisson.html.

[114] NERC. *Data collection for Demand-side management*. Tech. rep. North american electric reliability corporation, 2007.

[115] Verband der Netzbetreiber. *TransmissionCode 2003 - Anhang D1*. 2003. URL: www.regelleistung.net/.

[116] Nordel Demand Response Group. *Enhancement of Demand Response : final status report*. Apr. 2008. URL: http://fourfact.se/images/uploads/Nordel-DR.pdf.

[117] NorthWest Power and conservation council. *Sixth Northwest Conservation and Electric Power Plan*. Tech. rep. Chapter 5 : Demand Response. NorthWest Power and conservation council, 2010.

[118] M. Paulus and F. Borggrefe. "The potential of demand-side management in energy-intensive industries for electricity markets in Germany". In: *Applied Energy* 88.2 (2011), pp. 432–441.

[119] Johannes Pfeifenberger and Kathleen Spees. *Characteristics of Successful Capacity Markets*. APEx Conference 2013 - New York, October 31. The Brattle Group. 2013. URL: http://www.brattle.com/.

[120] PJM interconnection. *Generation 301 : ancillary services. ancillary services*. 2013. URL: http://www.pjm.com/training.

[121] Serge Poignant and Bruno Sido. *Rapport Poignant - Sido. Groupe de travail sur la maitrise de la pointe électrique*. Rapport au ministre Jean-Louis Borloo. Apr. 2010.

[122] D. Rivola, A. Giusti, M. Salani, A.E. Rizzoli, R. Rudel, and L.M. Gambardella. "A decentralized approach to demand side load management: The Swiss2Grid project". In: *Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE*. Nov. 2013, pp. 4704–4709. DOI: 10.1109/IECON.2013.6699895.

[123] Robert G. Gallager. *Stochastic Processes - Theory for Applications*. Cambridge University Press, 2014. ISBN: 9781107039759.

[124] L. Ruff. *Economic principle of demand response in electricity*. Tech. rep. Edison Electric Institute, 2002.

[125] F.C. Schweppe. *Frequency adaptive, power-energy re-scheduler*. US Patent 4,317,049. Feb. 1982. URL: https://www.google.com/patents/US4317049.

[126] F.C. Schweppe, R.D. Tabors, J.L. Kirtley, H.R. Outhred, F.H. Pickel, and AJ. Cox. "Homeostatic Utility Control". In: *Power Apparatus and Systems, IEEE Transactions on* PAS-99.3 (May 1980), pp. 1151–1163. DOI: 10.1109/TPAS.1980.319745.

[127] J.A. Short, D.G. Infield, and L.L. Freris. "Stabilization of Grid Frequency Through Dynamic Demand Control". In: *Power Systems, IEEE Transactions on* 22.3 (Aug. 2007), pp. 1284–1293. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2007.901489.

[128] Ramteen Sioshansi. "Modeling the Impacts of Electricity Tariffs on Plug-in Hybrid Electric Vehicle Charging, Costs, and Emissions". In: *Operations Research* 60.3 (May 2012), pp. 506–516. DOI: 10.1287/opre.1120.1038. URL: http://ise.osu.edu/ISEFaculty/sioshansi/papers/PHEV_tariff.pdf.

[129] K. Smith and R. Hledik. *Drivers of Demand-response adoption : past, present and future*. The brattle group, Johnson Controls, Mar. 2011. URL: http://www.institutebe.com/InstituteBE/media/Library/Resources/Smart%20Grid_Smart%20Building/Issue-Brief---Demand-Response-Drivers,-ENG.pdf.

[130] Kathleen Spees and Lester B. Lave. "Demand Response and Electricity Market Efficiency". In: *The Electricity Journal* 20.3 (2007), pp. 69–85.

[131] I. Stadler. "Demand response: nichtelektrische Speicher fur Elektrizitatsversorgungssysteme mit hohem Anteil erneuerbarer Energien". 2006.

[132] Ingo Stadler. "Power grid balancing of energy systems with high renewable energy penetration by demand response". In: *Utilities Policy* 16.2 (2008), pp. 90–98.

[133] S. Stoft. *Power System Economics*. John Wiley and Sons, 2002.

[134] Jessica Stromback, Christophe Dromacque, and Yassin Mazin. *The potential of smart meter enabled programs to increase energy and systems efficiency : a mass pilot comparison - Empower Demand*. 2011. URL: http://esmig.eu/sites/default/files/2011.10.12_empower_demand_report_final.pdf.

[135] S. Sundaram and C.N. Hadjicostis. "Delayed Observers for Linear Systems With Unknown Inputs". In: *Automatic Control, IEEE Transactions on* 52.2 (Feb. 2007), pp. 334–339. ISSN: 0018-9286. DOI: 10.1109/TAC.2006.890371.

[136] Shreyas Sundaram. *Fault-Tolerant and Secure Control Systems - Course Notes*. 2012. URL: https://ece.uwaterloo.ca/~ssundara/courses/notes/linear_systems.pdf.

[137] Emily Taylor. "The environmental impact of smart grids communication systems: a case-study approach". Master Thesis. Université Catholique de Louvain, 2015.

[138] Howard Taylor and Samuel Karlin. *An introduction to stochastic modeling*. Third. Academic Press, 1998.

[139] The Brattle group, Freeman-Sullivan & Co, and Global Energy Partners. *FERC Staff Report : A national Assessment of Demand Response Potential.* Tech. rep. Federal Energy Regulatory Commission, 2009. URL: `http://www.ferc.gov/legal/staff-reports/06-09-demand-response.pdf`.

[140] S.H. Tindemans, V. Trovato, and G. Strbac. "Frequency control using thermal loads under the proposed ENTSO-E Demand Connection Code". In: *PowerTech, 2015 IEEE Eindhoven.* June 2015, pp. 1–6. DOI: `10.1109/PTC.2015.7232676`.

[141] Jacopo Torriti, Mohamed G. Hassan, and Matthew Leach. "Demand response experience in Europe: Policies, programmes and implementation". In: *Energy* 35.4 (2010), pp. 1575–1583. DOI: `http://dx.doi.org/10.1016/j.energy.2009.05.021`.

[142] Transpower. *Automatic Under-Frequency Load Shedding (AUFLS).* Tech. rep. Transpower, 2010.

[143] Transpower. *System Security Forecast 2014. Part F - Dry Year Operational Issues.* Dec. 2014. URL: `www.systemoperator.co.nz/sites/default/files/bulk-upload/documents/Part%20F%20Dry%20Year.pdf`.

[144] Trevor A. Marshall. *Using Dishwashers to Provide Frequency Regulation Though Dynamic Demand.* Tech. rep. Graduate School of the Environment, Centre for Alternative Technology, 2011.

[145] UCTE. *Policy 1 Load-Frequency Control.* Tech. rep. UCTE, 2009.

[146] UK Office of Gas and Electricity Markets (Ofgem),Open Energi,KEMA. *The final report of the carbon emissions reduction target (CERT) 2008?2012.* 2013. URL: `https://www.ofgem.gov.uk/ofgem-publications/58431/npow08r12-118-report-081112.pdf`.

[147] Andreas Ulbig, Theodor S Borsche, and G Andersson. "Impact of low rotational inertia on power system stability and operation". In: *arXiv preprint arXiv:1312.6435* (2013).

[148] Deutsche ÜNB. *Eckpunkte und Freiheitsgrade bei Erbringung von Primärregelleistung.* 2014. URL: `https://www.regelleistung.net/ext/download/eckpunktePRL`.

[149] S. Vandael, T. Holvoet, G. Deconinck, S. Kamboj, and W. Kempton. "A comparison of two GIV mechanisms for providing ancillary services at the University of Delaware". In: *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on.* Oct. 2013, pp. 211–216. DOI: `10.1109/SmartGridComm.2013.6687959`.

[150] Rahul Walawalkar, Stephen Fernands, Netra Thakur, and Konda Reddy Chevva. “Evolution and current status of demand response (DR) in electricity markets: Insights from PJM and NYISO”. In: *Energy* 35.4 (2010), pp. 1553–1560.

[151] Shih-Ho Wang and E. Davison. “On the stabilization of decentralized control systems”. In: *Automatic Control, IEEE Transactions on* 18.5 (Oct. 1973), pp. 473–478. ISSN: 0018-9286. DOI: 10.1109/TAC.1973.1100362.

[152] S. Weckx, J. Driesen, and R. D’Hulst. “Optimal frequency support by dynamic demand”. In: *PowerTech (POWERTECH), 2013 IEEE Grenoble*. June 2013, pp. 1–6. DOI: 10.1109/PTC.2013.6652177.

[153] Wikler, Gosh, Marrin, Shishido, Carrera, and Faruqui. *The Potential for Demand Response to Integrate Variable Energy Resources with the Grid*. 2013. URL: http://www.brattle.com/.

[154] Matthias Winkel. *Part B Applied Probability*. 2007. URL: http://www.maths.ox.ac.uk/system/files/coursematerial/2012/2401/5/bs3a07.pdf.

[155] C.K. Woo, E. Kollman, R. Orans, S. Price, and B. Horii. “Now that California has AMI, what can the state do with it?” In: *Energy Policy* 36.4 (2008), pp. 1366–1374. ISSN: 0301-4215. DOI: http://dx.doi.org/10.1016/j.enpol.2007.12.018. URL: http://www.sciencedirect.com/science/article/pii/S0301421507005733.

[156] Z. Xu, J. Ostergaard, and M. Togeby. “Demand as frequency controlled reserve”. In: *Power Systems, IEEE Transactions on* 26.3 (2011), pp. 1062–1071.

[157] Jay W. Zarnikau. “Demand participation in the restructured Electric Reliability Council of Texas market”. In: *Energy* 35.4 (2010), pp. 1536–1543. ISSN: 0360-5442. DOI: http://dx.doi.org/10.1016/j.energy.2009.03.018. URL: http://www.sciencedirect.com/science/article/pii/S0360544209000565.

[158] Wei Zhang, Jianming Lian, Chin-Yao Chang, and K. Kalsi. “Aggregated Modeling and Control of Air Conditioning Loads for Demand Response”. In: *Power Systems, IEEE Transactions on* 28.4 (Nov. 2013), pp. 4655–4664. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2013.2266121.

[159] Charalampos Ziras, Evangelos Vrettos, and Goran Andersson. “Primary frequency control with refrigerators under startup dynamics and lockout constraints”. In: *Power Energy Society General Meeting, 2015 IEEE*. July 2015, pp. 1–5. DOI: 10.1109/PESGM.2015.7286291.

Appendix A

Price-based DR and decreased price risk.

We may illustrate the decreased price-risk consecutive to the introduction of price-based DR with a toy example. The own-price elasticity $\epsilon_{h,h}$ of demand at hour h is defined as the relative change in consumption $(q \rightarrow q')$ following a relative change in price $(p \rightarrow p')$.

$$\epsilon_{h,h} = \frac{q'_h}{q_h} \frac{p_h}{p'_h} \tag{A.1}$$

This elasticity has been evaluated in many experiments worldwide, which results have been complied by De Jonghe in [38]. It can be roughly approximated to be equal to -10%. Let's suppose a high price period in which the wholesale price reaches 200€/MWh. The demand at that time is $q_h =$ 10MWh. In a first situation, consumers are isolated from the price spike. Their contract offers them a flat tariff for electricity of $p_h = 100$€/MWh. In a second setting, the clients have decided on a pass-through contract with their supplier and are directly exposed to the wholesale price $p'_h = 200e/MWh$. With an own-price elasticity of -10%, the consumer will consume less than the initial 10MWh: $q'_h = 8$ MWh.

As a results of this process, the consumer's bill will increase of 600€ while its supplier's profit will pass from -2000€ in the initial situation to 0€. Exposing consumers to the wholesale prices allows suppliers to avoid losses. The rise in the energy bill should be compensated by lower off-peak prices.

Appendix B

Deriving the expected demand

The expected demand in equation (4.18) is derived below, using $S(t) = \lambda$ and $P = 1kW$.

- For time $t < T$, given that no load was started on $t < 0$,

$$\hat{D}(t) = \lambda P \int_0^t \int_\tau^{T_M} \delta(\tau - T)\, dT\, d\tau = \lambda P \int_0^t d\tau = \lambda P t$$

- For time $T \le t < t_1$, as the profile is constant

$$\hat{D}(t) = \lambda P T \tag{B.1}$$

- For time $0 \le t - t_1 < T_m$, all loads having started at a delayed instant $\tau > t - t_1$ have constant parameter ($\hat{D}_c(t)$). All new loads have varying parameter ($\hat{D}_v(t)$). And since $t - t_1 < T_m$, the distribution $\Omega_T(t)$ still plays no role in determining $\hat{D}(t)$.

$$\hat{D}(t) = \hat{D}_c(t) + \hat{D}_v(t) \tag{B.2}$$

$$\hat{D}_c(t) = \lambda P \int_{t-t_1}^{T} d\tau = \lambda P(T - (t - t_1)) \tag{B.3}$$

$$\hat{D}_v(t) = \lambda P \int_0^{t-t_1} \int_{T_m}^{T_M} \frac{1}{T_M - T_m}\, dT\, d\tau \tag{B.4}$$

$$= \lambda P(t - t_1) \tag{B.5}$$

$$\hat{D}(t) = \lambda P T \tag{B.6}$$

- For time $T_m \le t - t_1 < T$, a similar development is found, that takes into account the distribution $\Omega_T(t)$.

$$\hat{D}_c(t) = \lambda P \int_{t-t_1}^{T} d\tau = \lambda P(T - (t - t_1)) \tag{B.7}$$

$$\hat{D}_v(t) = \lambda P \left(T_m + \int_{T_m}^{t-t_1} \int_\tau^{T_M} \frac{1}{T_M - T_m}\, dT\, d\tau \right) \tag{B.8}$$

$$= \lambda P(T_m + \frac{2T_M(t - t_1 - T_m) - (t - t_1)^2 + T_m^2}{2(T_M - T_m)}) \tag{B.9}$$

$$\hat{D}(t) = \lambda P(T - (t - t_1) + T_m + \frac{2T_M(t - t_1 - T_m) - (t - t_1)^2 + T_m^2}{2(T_M - T_m)}) \tag{B.10}$$

- For time $T \leq t - t_1 < T_M$, there are no more loads with constant parameter, while the probability for having loads with duration in the non-empty interval $[t - t_1, T_M]$ is zero.

$$\hat{D}(t) = \lambda P(T_m + \frac{2T_M(t - t_1 - T_m) - (t - t_1)^2 + T_m^2}{2(T_M - T_m)}) \quad \text{(B.11)}$$

- When $T_M \leq t - t_1 < t_2 - t_1$, the demand recovers its equilibrium expected demand, which in this case corresponds to the initial situation.

$$\hat{D}(t) = \lambda P \frac{T_M + T_m}{2} = \lambda PT \quad \text{(B.12)}$$

Appendix C

Approximation of the Aggregate model in the Rate policy

Let's recall the three aggregate models relative to each of the presented control policies,

$$x^d(t|T_n) = \tilde{\alpha}_{T_n}(t) - \tilde{\alpha}_{T_n}(t - T_n) \tag{C.1}$$
$$x^s(t|T_n) = \alpha(t) \quad - \bar{\alpha}(t, T_n) \tag{C.2}$$
$$x^r(t|T_n) = \alpha(t) \quad - \big(1 + \alpha(t)\big)n^r(t)\bar{\alpha}(t, n^r(t)T_n) \tag{C.3}$$

We would like to find a linear approximation of the non-linear rebound of the Rate policy. We will show that, as the frequency-based reference signal has a small energy content (i.e., $\int_0^{T_n} r^{up}(t-\tau)d\tau$), the following approximation is valid (equations (C.3) and (C.2)) for large T_n.

$$x^r(t) \simeq x^s(t) \qquad \Rightarrow \qquad \bar{\alpha}(t, T_n) \simeq \big(1 + \alpha(t)\big)n^r(t)\bar{\alpha}(t, n^r(t)T_n) \tag{C.4}$$

We want to evaluate the approximation error $e_a(t)$.

$$e_a(t) = x^r(t) - x^s(t) \tag{C.5}$$

We pose for readability and feasibility $u(t) = (1+K_D r^{up}(t))/T_n > 0$ (i.e., $0 \le K_D < 1$ without loss of generality) and recall that $\alpha(t) = K_D r^{up}(t)$. We use equation (5.41) as the definition of the dynamic run time.

$$1 \quad = \int_{t-T^{dyn}(t)}^{t} u(s)\, ds \tag{C.6}$$
$$e_a(t) = \frac{1}{T_n}\left[\int_{t-T_n}^{t} \alpha(s)\, ds - (1+\alpha(t))\int_{t-T^{dyn}(t)}^{t} \alpha(s)\, ds\right] \tag{C.7}$$

where $T^{dyn}(0) = T_0$ depends on $u_0 = \{u(t)|t \le 0\}$ is known (e.g., $T_0 = T_n$ if $u_0 = 1$).

Dynamic Run Time

Let's denote $\tau(t) = T^{dyn}(t)$ and let's take the time derivative of the first equation.

$$\frac{d\tau(t)}{dt} = 1 - \frac{u(t)}{u(t-\tau(t))} \tag{C.8}$$

Equation (C.8) is a non-linear delay differential equation with time varying delay. It is dependent on the initial delay value $\tau(t \leq 0) = \tau_0$. As illustrated on figure C.1, it is an infinite dimension problem as the future evolution of $\tau(t)$ are dependent on all values of $u(t)$ in the interval $[t-\tau(t), t]$ [109]. These values form a function u_t.

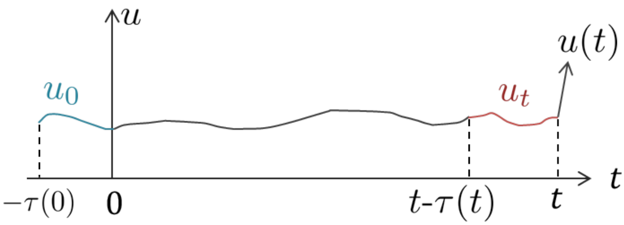


Figure C.1: The evolution of the time delay $\tau(t)$ is function of all previous values u_t.

As the initial reference signal $r^{up}(t)$ is bounded, the time delay is bounded also bounded to $T_n \leq \tau(t) \leq \tau_M = T_{dl} = T_n/(1-K_D)$. The relation between the previous and current values of the delay are described in a functional which domain is a space of functions. Those functions are defined on the domain $[-\tau_M, 0]$ and codomain $[-K_D, 0]$ (i.e., possible values of $u(t)$). The codomain of the functional is the set of all possible values of the delay $[-\tau_M, 0]$. The domain of the functional is the *Banach space of continuous functions over* $[-\tau_M, 0]$: $\mathcal{C}([-\tau_M, 0], \mathbb{R})$ [109]. The functional is therefore described by $\mathcal{C}([-\tau_M, 0], [-K_D, 0]) \rightarrow [-\tau_M, 0]$. As $u(t)$ is greater than zero at all time, the time derivative of $\tau(t)$ exists at all point in time. Furthermore, $u(t)$ is most often close to $1/T_n$, which leads $\tau(t)$ to be close to T_n, and $e_a(t)$ to be close to zero.

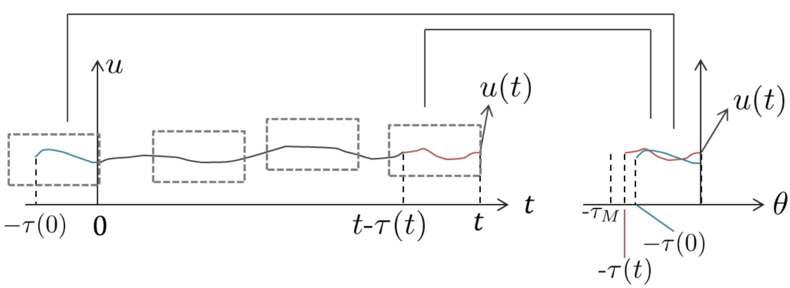


Figure C.2: A sliding window view on evolution of the time delay $\tau(t)$.

As can be observed on figure C.2, the required information can be condensed in a sliding window of length τ_M. The new information comes exclusively from the right (i.e., $u(t)$). Indeed, the required information lies at t in the interval $[t-\tau(t), t]$. After an infinitesimal time dt, the new information lies in $[t+dt-\tau(t+dt), t+dt]$. We would like to show that the new rightmost limit $t+dt-\tau(t+dt)$ is always greater than the initial limit $t-\tau(t)$ such that no new information was needed from the past.

$$\lim_{dt\to 0} t+dt-\tau(t+dt) > t-\tau(t) \tag{C.9}$$

This is always true as it gives

$$\lim_{dt\to 0} \frac{\tau(t+dt)-\tau(t)}{dt} < 1 \tag{C.10}$$

$$\frac{d\tau(t)}{dt} < 1 \tag{C.11}$$

which from equation (C.8) with $u(t) > 0$ is indeed true.

Approximation Error

When the delay $\tau(t)$ is known, the error can be computed independently. Let's use $I_\alpha(t)$ as the integrated $\alpha(t)$ signal.

$$e_a(t) = \frac{1}{T_n}\Big[I_\alpha(t) - I_\alpha(t-T_n) - \big(1+\alpha(t)\big)\big(I_\alpha(t) - I_\alpha(t-\tau(t))\big)\Big] \tag{C.12}$$

This error is illustrated for $K_D = 60\%$ and both $T_n = 1/4h$ or $T_n = 1h$ on figure C.3. The error decreases largely as T_n increases.

The maximum observed error as simulated from January 2015 data is represented below on figure C.4.

Upper bound error

We can also derive an upper bound for the error term.

$$|e_a(t)| \leq \frac{1}{T_n}\Big|I_\alpha(t-\tau(t)) - I_\alpha(t-T_n)\big)\Big| \tag{C.13}$$

By noting $z(t) = \tau(t) - T_n$, the upper bound of $e_a(t)$ is

$$|e_a(t)| \leq \max_t \frac{1}{T_n}\int_0^{z(t)} |\alpha(t-T_n-s)|\, ds \tag{C.14}$$

The solution can be found by using specific DDEs solvers to compute, at any time t, the delay $\tau(t)$ and thereby being able to compute the error $e_a(t)$.

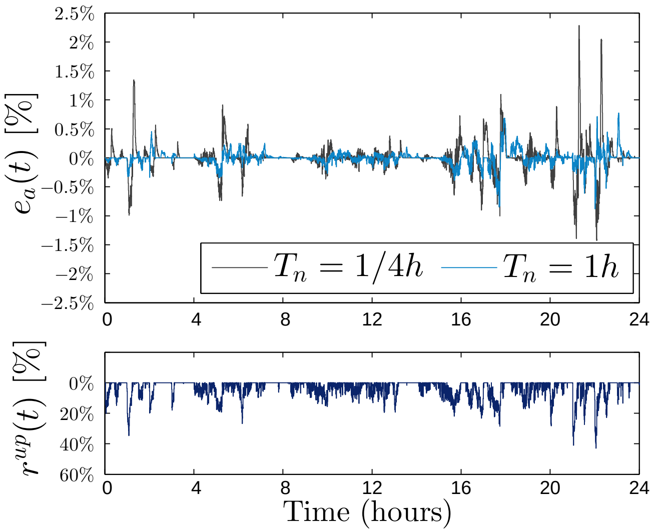


Figure C.3: Illustration of the approximation error magnitude $e_a(t)$ in % of the baseline demand. The error is a function of the reference signal $r^{up}(t)$, shown on the bottom chart (1st January 2015, source : RTE). Parameters : $K_D = 60\%$ and $T_n = 1/4h$ or $T_n = 1h$.

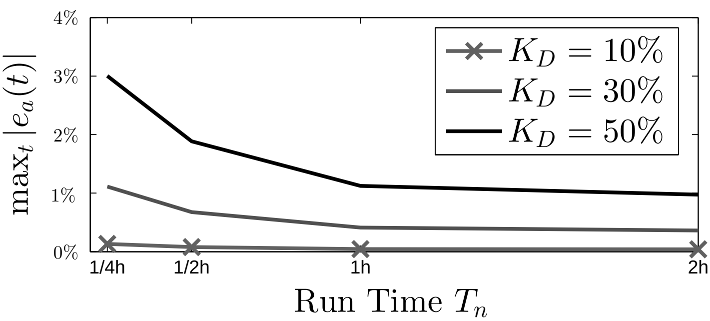


Figure C.4: Maximum observed approximation error magnitude $e_a(t)$ in % of the baseline demand (January 2015, source : RTE) for different parameters T_n and K_D.

Discrete-time model

DDEs solver can be quite slow to converge. Therefore, an equivalent discrete-time approach can be used. At every time step k of length T_s, a large matrix is used to keep track of the amount of energy that remains to be consumed by each running loads. The discrete approach considers that $S(k) = \lambda T_s$ loads start at each time step k. At every time step, an amount of energy $\Delta E(k) = P(k)T_s$ is consumed by the loads at a rate $P(k) = P_n(1 + b(k))$. The bias $b_k = K_D r_k^{up}$ is given by the average value $r^{up}(k)$ of $r^{up}(t)$ in the interval $t \in [T_s k, T_s(k + 1)[$. Also, each running load is associated to two energy states $E^b(k)$ and $E^e(k)$. They both account for the amount of energy that still need to be consumed before the load can stop, evaluated respectively at the beginning (b) and at the end (e) of period k. They must always stay positive, which requires a non-convex operation (binary-like). All loads starting within a certain time step are associated to the same energy states. This reduces largely the computational burden as the number of variable shrinks to the largest relevant time window : $T_n/T_s/(1 - K_D)$ (or the next integer time step).

For instance, we define $E^{b,i}(k)$ as the remaining energy of loads being started since i time steps at time k, $\forall i = 0..I$, with $I = \lceil T_n/T_s/(1 - K_D)\rceil$. Such states are successively computed to track past efforts $P(k)$. Demand at time k is computed as follows.

$$
\begin{aligned}
D(k) &= \lambda T_s(\sum_{i=0}^{I} E^{b,i}(k) - E^{e,i}(k))/T_s && \text{(C.15)}\\
E^{e,i}(k) &= \max(E^{b,i}(k) - P(k), 0) && \text{(C.16)}\\
E^{b,0}(k + 1) &= E_n && \text{(C.17)}\\
E^{b,i}(k + 1) &= E^{e,i-1}(k) \quad \forall i > 0 && \text{(C.18)}
\end{aligned}
$$

The error is simply extracted from its definition (Eq. C.5).

$$
\begin{aligned}
x^r(k) &= \frac{D(k)}{D_0} - 1 && \text{(C.19)}\\
x^s(k) &= b(k) - \frac{T_s}{T_n}\sum_{j=1}^{T_n/T_s} b(j) && \text{(C.20)}\\
e_a(k) &= x^s(k) - x^r(k) && \text{(C.21)}
\end{aligned}
$$

Appendix D

Imbalance recovery for simulation purpose

D.1 Immediate damping

The power system is assumed to have a constant damping factor $K_{FCR}+L$ that is immediately deployed as soon as a frequency deviation is observed. The assumed power system model is reduced to its minimum. The resulting imbalance $I_1(t)$ aggregates together the actual system imbalance but also all operator's effort to counteract this imbalance (e.g., aFFR and mFRR volumes).

$$I_1(t) = (y_f(t) - f_n)(K_{FCR} + L) \tag{D.1}$$

The system described by equation (6.1) must be reduced to its minimum where K_{aFRR} and D_0 are set to zero. Indeed, the contribution of the integral term is already included into $I_1(t)$. Note that in this case, the finite ramp rate of generators is not considered to build the input $I_1(t)$. Furthermore, the historical integral action cannot be recovered. The model used in simulations will be as follows (Eq. (D.2)). In order to recovering $I_1(t)$, the system below is considered without load participation (i.e., $D_0 = 0$, $K^{dn}_{FCR} = K^{up}_{FCR} = 15GW/Hz$).

$$M\frac{df_1(t)}{dt} = I_1(t) - L\Delta f_1(t) - K^{dn}_{FCR}\Delta f_1^+(t) - K^{up}_{FCR}\Delta f_1^-(t) - D_0 x_1(t) \tag{D.2}$$

The resulting $f_1(t)$ are a little different, but very close, to the initial frequency measurements. The main limitation of this approach is that the quality of the frequency $f_1(t)$, when demand response is introduced ($D_0 > 0$ and $K_D > 0$), will be very poor. Indeed, the rebound will not be compensated by $aFRR$ as shown in figure 6.7. Indeed, there is no integral action considered in this case. We anticipate on the results presented below to show such effect (Fig. D.1). The two presented histograms represent respectively the historical frequency quality, and the simulated frequency quality with load participation. In case flexible load participation is considered, the groups covers the whole upward FCR volumes. Ideal generation still covers the downward volumes. The histograms are presented for the different group scenarios/aggregate models and compared to the input data. As highlighted, the frequency quality is decreased when no integral action is considered to counteract the rebound error.

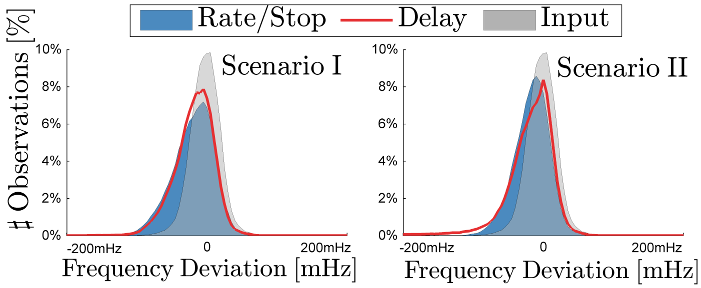


Figure D.1: Frequency deviation histogram as image of general frequency quality. Input Data are from January 2015 (source : RTE). Grey area : input data. Blue area : Stop/Rate aggregate model. Red line : Delay aggregate model.

D.2 Immediate damping with integral compensation

In order to compensate for the rebound, and uniquely for the rebound, we construct an additional $aFRR$ layer supposed to keep the frequency quality as simulated without demand response. We use the input data as reference for this additional layer. These slow reserves will attempt to bring back frequency back to its historical value. Let's note that the imbalance is such that $I_2(t) = I_1(t)$.

$$M\frac{df_2(t)}{dt} = I_2(t) - L\Delta f_2(t) - K_{FCR}^{dn}\Delta f_2^+(t) - K_{FCR}^{up}\Delta f_2^-(t) \\ - \frac{K_{aFRR}}{\tau_{FRR}}\int (f_2(t) - y_f(t))dt - D_0 x_2(t) \tag{D.3}$$

D.3 l-delay unknown input observer : finite ramp rate of generators

An observer can be built to recover the unknown input [135]. The basic idea is to reconstruct all observable states and input at a certain time t from the l previous output measurement. This approach allows to consider more complex initial systems. Let's suppose the linear discrete time ABCD system with states $\mathbf{x} = (x_1, x_2)$ and input $u \in \mathbb{R}$.

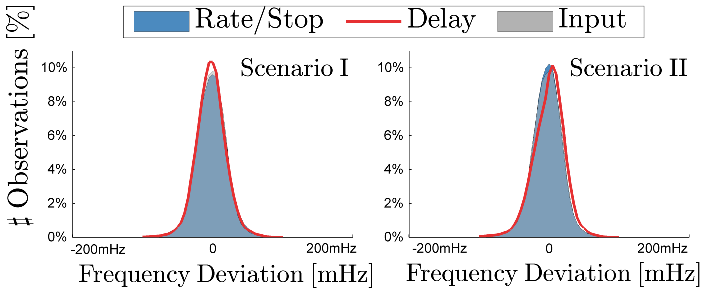


Figure D.2: Frequency deviation histogram (with integral compensation).

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + Bu(t) \tag{D.4}$$
$$y(t) = Cx(t) + Du(t) \tag{D.5}$$

As explained in [136], the system may be iterated from an initial value $\mathbf{x}(0)$.

$$\underbrace{\begin{pmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(l) \end{pmatrix}}_{y(0:l)} = \underbrace{\begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^l \end{pmatrix}}_{\mathcal{O}_l} \mathbf{x}(0) + \underbrace{\begin{pmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{l-1}B & CA^{l-2}B & \cdots & D \end{pmatrix}}_{\mathcal{J}_l} \underbrace{\begin{pmatrix} u(0) \\ u(1) \\ u(2) \\ \vdots \\ u(l) \end{pmatrix}}_{u(0:l)} \tag{D.6}$$

This may be rewritten in a compact form.

$$y(0:l) - \mathcal{O}_l\mathbf{x}(0) = \mathcal{J}_l u(0:l) \tag{D.7}$$

Our objective is to invert the above relation in order to obtain the unknown inputs $u(0)$,$u(1)$,etc. Under several conditions, the above system is said to be *invertible*. As mentioned in [136] :

> The system is said to have an l-delay inverse if it is possible to uniquely recover the input $u(t)$ from the outputs of the system up to time-step $y(t+l)$ (for some nonnegative integer l), assuming that the initial state $x(0)$ is known.

The required conditions are that, for a system with m inputs, the first m columns of $\mathcal{J}_l$ are both linearly independent of each other and independent of all *other* columns of $\mathcal{J}_l$. The idea is to find a delay $l \leq n - \text{nullity}(D) + 1$ (i.e. n is the system's dimension) such that,

$$\text{rank}[\mathcal{J}_{l+1}] = m + \text{rank}[\mathcal{J}_l] \qquad \forall 1 \leq l \leq n - \text{nullity}(D) + 1 \tag{D.8}$$

In the case of equation 6.1, we have one unknown input ($u(t) = I(t)$), and two states $x_1(t) = \Delta f(t)$, $x_2(t) = \int \Delta f(t)$. The condition boils down to insuring that we can find a delay l that respects eq. (D.8). In such case, it is possible to find a matrix P such that $P\mathcal{J}_l$ degenerates into a vector $P\mathcal{J}_l = [1\ 0\ ...\ 0]$. This gives,

$$P\big[y(0:l) - \mathcal{O}_l\mathbf{x}(0)\big] = u(0) = I_2(0) \tag{D.9}$$

Knowing $u(0)$ allows to compute $x(1)$ from equation (D.4) and re-iterate the procedure to obtain $u(1), u(2)$,etc. In fact, what appears to be necessary is to find estimates $\hat{\mathbf{x}}(1), \hat{\mathbf{x}}(2)$,etc. of the states $\mathbf{x}(1), \mathbf{x}(2)$,etc. from the initial value $\mathbf{x}(0)$ in order to successively recover the unknown input. We must build the following system.

$$\hat{\mathbf{x}}(t+1) = E\hat{\mathbf{x}}(t) + Fy(t:t+l) \tag{D.10}$$

Again from [136], the system (D.10) is said to be an unknown input observer with delay l if the estimation error ($\hat{\mathbf{x}}(t) - \mathbf{x}(t)$) asymptotically converges to zero as t goes to infinity regardless of the value of $u(t)$. More details are to be found in appendix E.

The difference with the first approach is that the considered power system model can be more complex, and its parameters may in principle be time varying. This would allow to capture a much more realistic behavior. This approach however requires a good understanding of the initial system at stake. Indeed, a pole placement procedure is required in some cases order to select the gain of the observer. Furthermore, the procedure is not possible for all system parameters. The ABCD matrices must be have some desired features without which the approach is not converging.

The l-delay process is applied to system of equation (D.11) where generators that participates in FCR_G have finite ramp rate (first order model). Again, for recovering $I_3(t)$, D_0 is set to zero.

$$M\frac{df_3(t)}{dt} = I_3(t) - L\Delta f_3(t) + FCR_G(t) - D_0 x_3(t) \tag{D.11}$$

$$\tau_g\frac{FCR_G(t)}{dt} = -K_{FCR}^{dn}\Delta f_3^+(t) - K_{FCR}^{up}\Delta f_3^-(t) - FCR_G(t) \tag{D.12}$$

The figure D.3 below illustrates the obtained values of $I_1(t)$ (immediate damping) and $I_3(t)$. The difference are very small and consists in small transition periods that have a first order response shape. The added value of the above approach in such case is very small, partly due to the low refreshment rate of the initial data (10s-based measurements).

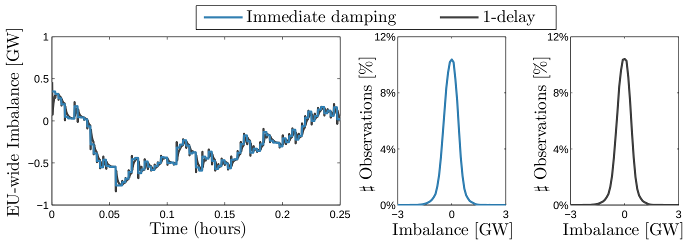


Figure D.3: Comparison between immediate damping and l-delay input reconstruction procedures. The power system model is considered without integral action but with limited ramp rate of generation assets.

D.4 l-delay unknown input observer for model with integral action.

The same procedure is applied to a complex model where the integral action is considered at early stage. In principle, such model allows to distinguish the initial system power imbalance from the consecutive action undertaken by the power system operators to push back frequency to its initial value are extracted from the imbalance signal.

$$M\frac{df_4(t)}{dt} = I_4(t) - L\Delta f_4(t) + FCR_G(t) + aFRR_G(t) - D_0 x_4(t) \quad \text{(D.13)}$$

$$\tau_g \frac{FCR_G(t)}{dt} = -K^{dn}_{FCR}\Delta f_4^+(t) - K^{up}_{FCR}\Delta f_4^-(t) - FCR_G(t) \quad \text{(D.14)}$$

$$\tau_{aFRR}\frac{aFRR_G(t)}{dt} = -K^p_{aFRR}\Delta f_4(t) - K^i_{aFRR}\int \Delta f_4(t)dt - aFRR_G(t) \quad \text{(D.15)}$$

As can be seen from the above equation, aFRR can, in general, provide a proportional (K^p) and an integral (K^i) action. Results of the imbalance $I_4(t)$ extracted for 1 day (1st January 2015) and 1 month (Jan. 2015) with a 1-delay observer are show respectively on Fig. D.4 and D.5 ($\tau_g = 3$,$\tau_{aFRR} = 5$,$K_{FCR} = 15GW/Hz$,$K^p_{aFRR} = 0$,$K^i_{aFRR} = 0.1K_{FCR}$,$L = 4GW/Hz$).

As can be observed the reconstructed input values rise to high values. Such imbalance value seems around 25 times higher than what could be expected at CE-level. Indeed, a simple parallel with the Belgian situation may help to assess this order of magnitude. We had access to the power

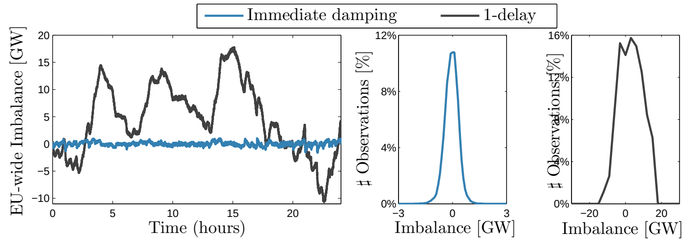


Figure D.4: Comparison of EU-wide imbalance recovered by two different methods (1) Immediate damping and (2) l-delay approach for model with integral action. Data from 1st of January 2015 (Source : RTE).

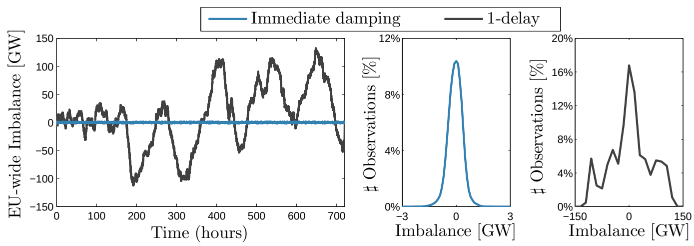


Figure D.5: Comparison of EU-wide imbalance recovered by two different methods (1) Immediate damping and (2) l-delay approach for model with integral action. Data from the month of January 2015 (Source : RTE).

imbalance signal as measured at the border of the Belgian System in January and September 2012. As highlighted on figure D.6, the instantaneous power system imbalance reaches at most 1.5 GW in the Belgian system, that has a peak load of approximately 12GW. Knowing that the peak load of the CE system is 400GW, and supposing that the imbalance increases as the square root of the peak load, we can estimate the CE-level imbalance to be around $1.5\sqrt{400/12} = 15/\sqrt{3} = 8.66$. This is indeed around 25 times less than the obtained imbalance signal $I_4(t)$. Let's note than even a proportional increase of the imbalance with the peak load would lead to 50GW, which is three times less than maximum of I_4.

D.5 Model selection conclusion

The complex model used to recover the imbalance signal without the integral action is inappropriate. Indeed, the non-zero frequency integral has other origins than the actual power imbalance. It is also linked to non-linear be-

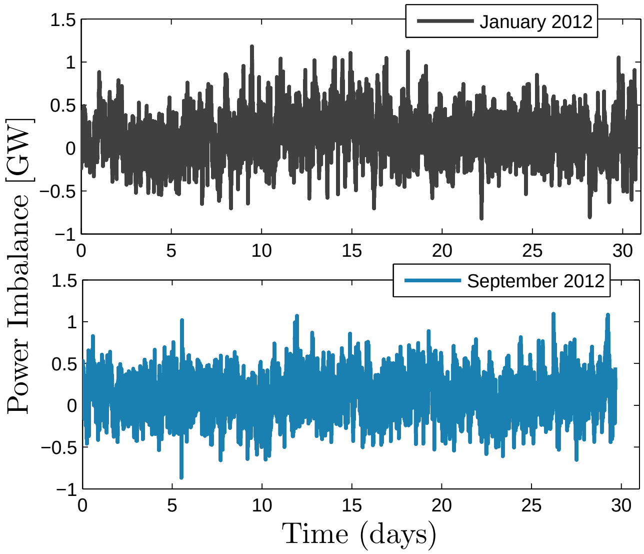


Figure D.6: Historical Power Imbalance measured at the border of the Belgian System during two different months in 2012 (10s-based measurements).

havior of the power-frequency control in general : saturation effects, limited ramp rates, dead-band, ACE netting, measurement errors, ACE resetting at local level, etc. Recovering the integral action seems out of reach as all of such elements should be taken into account. We therefore decide to use the second model (immediate damping with integral action for rebound management).

Appendix E

Design of unknown input observer

From [136], we learn that the following observer is said to be a L-delay unknown input observer if the estimation error of the states $\hat{\mathbf{x}}(t) - \mathbf{x}(t)$ is asymptotically null as t goes to infinity. We disposed of historically measured inputs $y(t)$.

$$\hat{\mathbf{x}}(t+1) = E\hat{\mathbf{x}}(t) + Fy(t : t+L) \tag{E.1}$$

The design process required to find E and F is summarized in the following code.

Listing E.1: L-delay observer

```
function [E,F,l]=Ldelay(A,B,C,D,lmax)
% Construct the l-delay observer from system A,B,C,D.

[nx,nu]  = size(B);
[ny,nu2] = size(D);

% Invertibility and Observability matrices
Jl = D; Ol = C; Mat = D;
% recursion
k=1;
while k<=lmax

    % Save for later
    Jlm1 = Jl;
    bm=size([Ol*B Jl],2);

    Jl = [D zeros(ny,bm-nu2);Ol*B Jl];
    Ol = [C; Ol*A];

    Mat = [Mat; C*A^(k-1)*B];

    % Stop if the delay is sufficient
    bm1=size([Ol*B Jl],2);
    JLp1 = [D zeros(ny,bm1-nu2);Ol*B Jl];
    if rank(JLp1)- rank(Jl) == nu
        l = k;
        k=lmax+1;
```

```
    elseif k == lmax
        error('Full State cannot be recovered')
    else
        k=k+1;
    end
end

% Nbar matrix
Nbar = null(Jlm1')';

% W matrix
[an,bn] = size(Nbar);
Mat2 = [eye(ny) zeros(ny,bn) ; zeros(an,ny) Nbar];
Mat3 = Mat2*Mat;
Wup = null(Mat3')';
Wdn = ((Mat3')*Mat3)\(Mat3');
W=[Wup;Wdn];

% N matrix
N = W*Mat2;

% S1 and S2
Mat4 = N*Ol;
S1 = Mat4(1:end-nu,:);
S2 = Mat4(end-nu+1:end,:);

% F1 (place poles if needed)
if max(abs(eig(A-B*S2))) >= 1
    lambda = linspace(-0.5,0.5,nx);
    F1 = (place((A-B*S2)',S1',lambda)');
else
    as1=size(S1,1);
    F1=zeros(nx,as1);
end

% Generate outputs
E = A-B*S2-F1*S1;
F = [F1 B]*N;

end
```