Demand Response from Thermostatically Controlled Loads: Modelling, Control and System-Level Value

A thesis submitted to

Imperial College London

For the Degree of Doctor of Philosophy

by

Vincenzo Trovato

Department of Electrical and Electronic Engineering

Imperial College London

February 2015

Declaration

I hereby declare that the content of this thesis is a product of my own work and other work is appropriately referenced.

London,

February 24th 2015

Vincenzo Trovato

Abstract

The research area of this thesis concerns the efficient and secure operation of the future low-carbon power system, where alternative sources of control and flexibility will progressively replace the traditional providers of ancillary services i.e. conventional generators. Various options are engaged in this challenge and suit the innovative concept of Smart Grid. Specifically, this thesis investigates the potential of demand side response support by means of thermostatically controlled loads (TCLs).

This thesis aims to quantify the impact that a population of thermostatically controlled loads has on the commitment and dispatch of a future power system characterized by a large penetration of renewable energy sources (e.g. wind) that are variable and intermittent. Thanks to their relative insensitivity to temperature fluctuations, thermostatic loads would be able to provide frequency response services and other forms of system services, such as energy arbitrage and congestion relief. These actions in turn enhance the power system operation and support the strict compliance with system security standards.

However, the achievement of this transition requires addressing two challenges. The first deals with the design of accurate device models. Significant differences affect the devices' design included in the same class, leading to different system-level performances. In addition, the flexibility associated to TCLs would be handled more easily by means of models that describes the TCLs dynamics directly as a cluster rather than considering the appliances individually. Second, it is not straightforward achieving satisfactory controllability of a cluster of TCLs for the considered applications. The complexity lies in the typical operation of these devices that has only two power states (*on* and *off*) whereas the desired response is continuous. Moreover the control strategy has always to comply with strict device-level temperature constraints as the provision of ancillary services cannot affect the quality of the service of the primary function of TCLs.

This thesis addresses the challenges exhibited. Detailed thermal dynamic models are derived for eight classes of domestic and commercial refrigeration units. In addition, a heterogeneous population of TCLs is modelled as a *leaky storage unit*; this unit describes the aggregate flexibility of a large population of TCLs as a single storage unit incorporating the devices' physical thermal models and their operational temperature limits. The control problem is solved by means of an initial hybrid controller for frequency response purposes that is afterwards replaced by an advanced controller for various applications. Provided these two elements, a novel demand side response model is designed considering the simultaneous provision of a number of system services and taking into account the effect of the load energy recovery. The model, included in a stochastic scheduling routine, quantifies the system-level operational cost and wind curtailment savings enabled by the TCLs' support.

Publications including material of the thesis

Journal Papers:

S. Tindemans, V. Trovato and G. Strbac: *"Decentralised control of thermostatic loads for flexible demand response"* –IEEE Transactions on Control Systems Technology (accepted Nov. 2014).

V. Trovato, S. Tindemans and G. Strbac: *"The Leaky Storage Model for Optimal Multi-service Allocation of Thermostatic Loads"* – submitted to IET Generation, Transmission & Distribution (Jen. 2015).

F. Teng, V. Trovato and G. Strbac: *"Stochastic Scheduling with Inertia-dependent Frequency Regulation"* – submitted to IEEE Transactions on Power Systems (Nov. 2014).

V. Trovato, F. Teng and G. Strbac: *"Stochastic Scheduling with Flexible Demand Response"* – to be submitted to IEEE Transactions on Smart Grid

S. De Boeck, K. Das, V. Trovato, J. Turunen, M. Halat, P. Sorensen and D. Van Hartem: *"Review of Defence Plans in Europe: Current Status, Strengths and Opportunities"* – IET Generation, Transmission & Distribution (Feb. 2015).

Conference Papers:

V. Trovato, S. Tindemans and G. Strbac: *"Controlling the Synchronization and the Payback Associated with the Provision of Frequency Services by Dynamic Demand"* – 22nd International Conference on Electricity Distribution, Stockholm, Sweden 10-13 June 2013.

V. Trovato, S. Tindemans and G. Strbac: *"Demand Response Contribution to Effective Inertia for System Security in the GB 2020 Gone Green Scenario"* – 4th IEEE-PES Innovative Smart Grid Technologies Europe (ISGT), Lyngby, Denmark 6-8 October 2013.

V. Trovato, S. Tindemans and G. Strbac: *"Security Constrained Economic Dispatch with Flexible Thermostatic Loads"* – 5th IEEE-PES Innovative Smart Grid Technologies Europe (ISGT), Istanbul, Turkey 12-15 October 2014.

V. Trovato, S. Tindemans and G. Strbac: *"Designing effective frequency response patterns for flexible thermostatic loads"* – submitted for presentation at IEEE-PES General Meeting, Denver, USA 26-30 July 2015.

S. Tindemans, V. Trovato and G. Strbac: *"Frequency control using thermal loads under the proposed ENTSO-E Demand Connection Code"* – submitted for presentation at Powertech conference, Eindhoven, The Netherlands 29 June-2 July 2015.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

TCL	*Thermostatically Controlled Load*
EU	*European Union*
NG	*National Grid*
GB	*Great Britain*
UK	*United Kingdom*
RES	*Renewable Energy Sources*
EC	*European Commission*
TSO	*Transmission System Operator*
AC	*Alternating Current*
RoCoF	*Rate of Change of Frequency [Hz/s]*
LFDD	*Low Frequency Demand Disconnection*
DSR	*Demand Side Response*
ENTSO-E	*European Network Transmission System Operator for electricity*
SCED	*Security Constrained Economic Dispatch*
SUC	*Stochastic Unit Commitment*
DSRM	*Demand Side Response Model*
SQSS	*Security and Quality of Supply Standards*
OFGEM	*Office of Gas & Electricity Markets*

Acknowledgements

Karl Popper used to say: “All life is problem solving”. In the last years, I had to face some of them; however none of them has been approached without receiving direct or indirect help from those people that make my life happy, rarely complicated but always beautiful.

Hence, it is a pleasure for me to say thanks to Angela for having been the perfect friend… every day during the last fifteen years. Thanks to my father, my mother, my sister and my brother in law; I could not imagine a more thought-provoking family; I love you all.

The deep help and guidance from my supervisor Goran Strbac is very much appreciated; I am happy to thank Simon Tindemans for the effort made helping me and his genuine courtesy, daily provided. I am also grateful to National Grid for financially supporting my Ph.D. research. The comments of Gareth Taylor and David Angeli, the two examiners of this thesis, are very much welcome.

There are no words to say how much grateful I am with Cristina, Giuseppe, Raffaele, Francesco, Fabio, Pierluca, Francesca, Pierluigi, Mario, Michele, Maria and Angelo. We all thank the inventors of Skype and low-cost flying companies.

London offered me so many gifts; some of them are Claudia, Alessandro, Jose, Marco, Andrea, Javier, Antonio, Elena, Thulasi, Inma, Chris, Lorenzo, Giorgio, Michele, Ankur and all the other lovely people met.

Finally I would like to thank the rest of my big and warm family and in particular Giovanna, Valentina, Agnese and my lovely grandmother.

Chapter 1: Introduction

1.1 Thesis context

Climate change is a severe issue that encompasses several and wide sectors of the existence on this planet. The consequences associated to this problem deal with an overall increase of the temperatures all over the world, a substantial change in the rainfall pattern, an increase of the ocean and sea level and an augmented rate of occurrence of extreme events that push into a risky territory the habitat of billions of people. In fact the global temperature has increased of 0.8 °C during the last 150 years [1]. The most effective defence against climate change is the abatement of greenhouse gas emissions.

1.1.1 Worldwide actions

The danger associated to climate change has been recognized at a global level with a milestone represented by the Kyoto Protocol [2] (signed in 1997 and adopted in 2005); this joint document forces developed countries to respect emission reduction targets. A more crucial shared vision has been adopted after the signing of the Cancun Agreements at the COP16 [3]. The outcome of the Cancun Agreements represents the largest collective effort ever achieved to drastically reduce the greenhouse emissions in a mutually accountable way. In fact, governments agreed that emissions need to be reduced to limit the global temperature increase below 2 Celsius degrees.

In this vein, the international community will be actively involved to reach effective results at the COP21 Conference in Paris in 2015 [4]. It is hoped that all the countries involved, especially those nations that hold the largest volumes of greenhouse gases' emissions, will sign a productive agreement on climate change.

1.1.2 European Union actions

The EU is actively involved in preventing dangerous climate change; in particular a strong abatement of CO_2 emissions must be performed by 2020 [5]. The 28 Member States have to reduce overall greenhouse gas emissions by 20% compared to the 1990 levels; this objective is part of the five headline targets of the EU 2020 strategy [6].

The EU contribution to the climate change prevention continues with the design of a roadmap for moving to a competitive low-carbon economy by 2050. This promising roadmap suggests that EU Member States should diminish the noxious emissions to 80% below the levels achieved in 1990 [7].

1.1.3 United Kingdom actions

The UK Government has played an active role in order to limit greenhouse gas emissions through legally binding targets, both now and in the future. UK has been signed up to the Kyoto Protocol since 1997 and the Climate Change Act, passed in 2008, stated the steps to be taken in order to reduce by 26% the greenhouse gases emissions by 2026 (considering the 1990 baseline) [8].

1.1.4 Impact of greenhouse gases emission targets on the power system sector

The electricity sector is highly involved in the climate change challenge. It is worth pointing out that, in this section, the UK context is taken as example and the discussion deals with the global electricity network context. A recent report from the UK Department of Energy and Climate Change revealed that the energy supply sector represents 38% of the total carbon dioxide emissions, which in turn represent 82% of the total UK greenhouse gas emissions [9]. The reason of such a high level of noxious emissions is due to the traditional strong dependency of the electricity generation on fossil fuels. The achievement of the aforementioned emission targets will cause the power system sector to adopt drastic changes to the traditional way it operates. The GB power system and the electricity networks around the world must integrate, over the next decades, a large share of renewable and low-carbon generation sources.

In this vein, the 2008 European Commission (EC) directive 2009/28/EC [10] established a legally binding target for renewable energy sources (wind, solar, hydropower, biomass etc.) to cover

up to 20% of the total energy consumption of the EU Members Sates. The UK Government jointly with the National Grid (NG), the GB Transmission System Operator (TSO), adopted the aforementioned EC directive and set up the so-called Gone Green Scenario for the year 2020; as a result, the wind generation capacity should rise up to 25-30 GW (10 GW installed capacity in 2013) [11, 12]. Moreover, at the beginning of January 2015 the wind power in the UK consists in a total installed capacity of 12 GW (8GW on-shore and 4 GW off-shore) [13]. Such a deep penetration allows the UK to be the sixth largest country for installed wind generation, having overtaken France and Italy in 2012 [14]. This level of wind penetration is expected to further increase in order to meet even more binding emission targets for years 2030 [15] and 2050 [16]. Finally, the accomplishment of these requirements would also imply large investments in the power system sector. The traditional generation portfolio is drastically changing; hence, the integration of a large share of RES and the decommissioning of high-emitting conventional power plants will require adequate investments. A recent report (July 2014) from the Department of Energy and Climate Change (DECC) and HM Government quantifies the amount of such an investment: "*Of the infrastructure investment that has happened since 2010 – and of the infrastructure investment pipeline out to 2020 and beyond – energy leads. In fact, the UK's future energy infrastructure investments are larger than transport, water and communications combined. From the £45 billion invested in electricity generation and networks alone between 2010 and 2013, to the energy project pipeline estimated at over £200 billion, the economic impact is huge*".

Since 1990, UK carbon dioxide emissions have decreased by around 21% [9]; in parallel the penetration of RES has grown. Although this result also depends on efficiency improvements of conventional technologies, it is largely driven by an increased RES penetration. However, the expected further integration of RES may not be sufficient to actually meet the emissions reduction targets, or more correctly, the future penetration could be limited by issues related to the security of the power system.

In the UK, apart from biomass, wind represents the largest source of renewable energy [17]. Wind turbines, with the exception of fixed speed induction generator type, make use of high-power converters to be connected to the AC network [18]. This mechanical decoupling prevents wind farms from naturally deploying the kinetic energy stored in their turbine's rotating masses [19]; wind generators do not therefore support the system inertial response. Clearly, the higher the wind penetration, the larger the system inertia decline will be. Considering such a reduced-inertia

network, in the unlucky event of a generation loss, the compliance with frequency security standards could become significantly challenging. The initial RoCoF [Hz/s] would achieve the minimum (negative quantity in this case) value few instants after the generator failure occurs and it may exceed the security threshold enabling further disconnections of further generators equipped with RoCoF-sensitive protection schemes [20]. Moreover, the frequency nadir may drop below the security limit with the risk of triggering costly national defence plans e.g. the LFDD [21]. If this is the case, large shares of the national system demand will be automatically disconnected; this action is performed in ten different tranches according to the amplitude of the frequency deviation, up to 60% of the system demand [21]. Finally, a shortage of conventional generators in favour of wind farms also reduces the available governors' response that facilitates the frequency stabilization around a desired intermediate steady state value. Consequently, the risk of severe consequences for system security and reliability rises.

As the TSO is obliged to maintain high system reliability standards, the integration of large shares of RES implies the need for increased amount of various types of ancillary services. Following a traditional approach, the ancillary services would be mostly delivered by means of part-loaded generators in combination with fast standing plants (committing conventional generators also increases the level of inertia). However, dispatching generators part-loaded not only leads to higher operation costs due to a decreased efficiency level, but also generates higher noxious emissions. These issues may eventually compromise the ability of the system to integrate growing RES.

The integration of large shares of RES in the power system operation, without taking further actions to obtain large amounts of ancillary services, is not a feasible way towards the achievement of future reduction targets on greenhouse gas emissions. Hence, academic and industry research activities on power system seek for an answer to how to meet environmental challenges in the most economic and sustainable way whilst maintaining network security.

1.2 Motivation

The previous section provided the general context within which this thesis lies. In particular, it has pointed out the need for economically sustainable joint actions in order to provide the increased amount of ancillary services required for the integration of large shares of RES. Wind generation, besides not contributing to the system inertial response, is also characterized by an

inherent variability, intermittency and non-controllability. Transmission system operators all around the world face complex challenges to guarantee flexible and secure network control over different time scales under future low-carbon scenarios. Furthermore, the traditional use of conventional generators as essentially unique source of ancillary services is no longer a credible option for future scenarios.

1.2.1 The Smart Grid concept

The idea of resorting to alternative sources of flexibility for power system operation and security represents one aspect of the wider Smart Grid concept. According to a document of *SmartGrids ETP* [22] a Smart Grid is "an electricity network that can intelligently integrate the actions of all users connected to it - generators, consumers and those that do both - in order to efficiently deliver sustainable, economic and secure electricity supplies. A Smart Grid employs innovative products and services together with intelligent monitoring, control, communication, and self-healing technologies to:

- better facilitate the connection and operation of generators of all sizes and technologies;
- allow consumers to play a part in optimizing the operation of the system;
- provide consumers with greater information and choice of supply;
- significantly reduce the environmental impact of the whole electricity supply system;
- deliver enhanced levels of reliability and security of supply."

The listed aims may engage different technologies from both the generation and the demand side. We refer to the use of HVDC to provide, for instance, inertial and primary frequency control [23]; a variety of ancillary services for the short-term and long-term system control is offered by demand side technologies such as electrical vehicles (primary frequency control [24] and energy arbitrage [25]) and energy storage [26, 18]. Furthermore, academic and industry research is actively focused on control frameworks that enable inertial response from wind farms [27]. Finally, domestic and commercial demand side appliances (e.g. refrigerators, HVAC, dishwashers etc.) can play a fundamental role within the Smart Grid concept [28, 29].

1.2.2 Potential for Demand Side Response with domestic and commercial appliances

The effective implementation of the Smart Grid concept may be achieved only in presence of a harmonious cooperation between all the aforementioned actors; none of the technologies listed above will be able to individually address all the challenges that arise in a future low-carbon power system. However, this thesis investigates in depth only one of these technologies: refrigeration; the ability and the benefits of providing demand side ancillary services with domestic and commercial appliances represent the main topic of this work. Although the deployment of the demand flexibility potential is not a new concept, the role of the DSR has started to attract continuously increasing interest only during recent years, now that environmental and security concerns are shaping the design of power system development.

The refrigeration appliances considered in this thesis are part of the class of TCLs. The primary function of these devices, such as refrigerators, air conditioners, heat pumps etc., is maintaining the appliances' temperature around a target value; this is achieved by altering the power consumption between two states. Referring to a refrigeration unit, in the *on* phase it absorbs its maximum power and reduces the controlled temperature; conversely, during the *off* phase, the power absorbed is negligible and the temperature therefore increases. State jumps are governed by means of a thermostat that operates using temperature deadband; when one of the two temperature thresholds is reached, the thermostat triggers a state switch. The possibility to access only two power levels (0 *off*, 1 *on*) represents a peculiarity of these devices compared to other demand side response devices.

For TCLs that are controlled by using a temperature deadband, small fluctuations in temperature are acceptable as long as the target temperature is approximately maintained over time. This characteristic amounts to the possibility of shifting demand from one moment in time to another in response to the network's needs, without noticeably affecting the quality of service. As a great number of these devices is regularly connected to the grid, making a strategic use of the intrinsic collective flexibility offered by TCLs would enable huge potential benefits. These devices could be smartly used to provide, for instance, controllable demand side response services to enhance the system frequency control after a large generator outage. If TCLs are able to detect a frequency drop (by means of local measurement systems), they would rapidly (few seconds) decide to switch themselves off in order to reduce the generation-demand imbalance. The duration of this ancillary service will depend on the type of security service provided; it may go from few tens of

seconds (primary response) up to several minutes (secondary response). Moreover, the actual ability to provide these services highly depends on device-level temperature constraints and intrinsic thermal inertia of the TCLs; the studies presented in Chapter 5 tackle in detail these issues. The interest on this technology is not only due to the significant energy and power levels involved (domestic units plus commercial units) but also due to a quite homogenous distribution of these appliances across the power system. Moreover, compared to electric vehicles and storage units, most[1] of the TCLs are already installed; this technology therefore may lead the roll out of the DSR integration within the system operations.

The integration of thermostatically controlled loads within the Smart Grid concept may enable the following benefits:

- ***Security Benefits***: the ability to rapidly switch off a cluster of devices in response to a frequency event may enhance the system inertial-primary response increasing the system security.
- **Economic Benefits**: the provision of fast response services reduces the need for on line part-loaded generators, thus increasing the efficiency of the system. Besides the frequency response services, TCLs may be used to enhance the dispatch of generators or to temporally relieve transmission lines' congestions; if, at a particular hour, the generation mix is expensive or high emitting, these devices can shift their aggregate consumption to adjacent periods exploiting their thermal storage capacity. Providing ancillary services to the system and shifting the aggregate demand, if required, would decrease the system operation cost as result of a more efficient system operation. Moreover, the roll out of DSR may have a positive impact on the society by establishing a new profitable business sector.
- **Environmental Benefits**: apart from economic benefits, a more efficient system operation due to active TCL participation would allow higher levels of RES penetration; the integration of such generation technologies will reduce the greenhouse gases emissions of the electricity sector.

[1] Heat pumps represent one of the most interesting category among the TCLs; however nowadays the penetration of these devices is still substantially low due to the high electricity cost compared to other primary sources of heating i.e. gas [98].

1.3 Demand side response with thermostatically controlled loads: current practice

The idea of adjusting the aggregate power consumption of a large population of thermostatic loads to provide various system services is not new; however only during the recent years, academic and industry research has reached milestones and fundamental results, driven by the ever more increasing interest around this field. An accurate investigation on this promising demand side technology has to cover three aspects: the first is the modelling of the TLCs´ temperature dynamics; the second deals with the design of an appropriate control strategy that guarantees the actual deliverability of any demand response service and ensures the recovery to a steady condition eventually. The last point regards the impact of TCLs' flexibility on the generators' commitment and, in particular, the possible reduction of the system operational cost and increased wind utilization.

1.3.1 Dynamic models of thermostatically controlled loads

In this thesis, the thermostatic loads modelled are refrigeration loads considering both the domestic and commercial sector. They represent a substantial base load as they are connected to the grid at all times, making them suitable for providing system services around the clock. Although the overall potential demand side support provided by heat pumps and generic electric heating systems would be significant, nowadays only 5% of the GB households make use of electricity as source of heating [30]. Moreover, the heating system modelling must deal with wide differences in the households' construction; in particular, the level of insulation, which largely affects the temperature dynamics, could be significantly different depending on the house typology.

Thermostatic loads are modelled through dynamic thermal models where the state variables are temperatures. Several works in literature describe the temperature dynamics of domestic refrigeration units by means of a first order ODE [31, 28, 32, 33, 34]. The use of first order models has a number of advantages; for instance, it reduces the computational burden when computer simulations involve a large number of devices. However, the unique state variable aims to simultaneously describe both the temperature of the internal wall of the unit's cabinet (i.e. the one controlled by the thermostat's sensors) and the temperature of the food to chill. Although the purpose of a refrigeration unit is to keep the stored food around a set-point temperature, the measured and controlled variable is the temperature of the internal wall of the unit. More accurate

formulations are offered in [29, 35]; these references consider combined domestic fridge-freezers; the use of second order models allow to separate the dynamics of the fridge compartment from the freezer's dynamics. For the remainder of this work we will refer to the *order* of the thermal models (e.g. second or higher) as the number of state variables that are described in the system (see (3.3) in Chapter 3) and not to the order of each differential equation in the system, that is always one. However, the solutions proposed in [29, 35] still do not permit a clear separation between the food temperature dynamics and the temperature evolution of the internal walls of each compartment (where the temperature sensors are placed).

Regarding the commercial sector, the available references in the literature describe the device temperature evolutions by means of high-order models [36, 37, 38]; however, it is not straightforward to make these models suitable for demand side control frameworks or optimization tools. The high order of these models reflects the complex design of some commercial units (e.g. refrigerators in supermarket are equipped with several compressors). Other commercial refrigeration units (e.g. bottle cooler, freezers, multidecks) are characterized by simpler designs and thus could be still modelled by low-order dynamic models without a substantial lack of accuracy.

1.3.2 Accurate control of thermostatically controlled loads for demand side response applications

The second aspect to be considered is the design of satisfactory control strategies that allows TCLs to supply various ancillary services to the system. The first controller has been proposed by Schweppe in 1979, with the unique aim to enhance the system frequency control [39]. The idea is to make use of thermostatic appliances that are insensitive to relatively short-term temperature excursions, in order to create a distributed load shedding that quickly helps the network to balance generation and demand, thus limiting frequency deviations. The fundamentals of the control framework proposed in 1979 have been recently reconsidered since the interest around the demand side response started to grow. In fact, several authors [28, 40, 29] used a heuristic modification of the regular temperature-threshold thermostatic controller that governs a TCL. This implementation has the advantage of simplicity and it is able to provide a clear support to the primary frequency control during the first instants after the generator outage; however, a more detailed investigation points out the inability to control the TCLs' long term response with this method. The appliances show in fact the tendency to synchronize their operating cycles; this effect

implies long-term fluctuations of the aggregate power consumption that may be translated into fluctuations of frequency.

The side effect of the synchronization is cancelled by means of the stochastic control strategy proposed in [31]. The switching from the *on* phase to the *off* phase and vice versa are stochastically determined and are not imposed by deterministic temperature boundaries anymore; this characteristic allows to cancel the synchronization. However, the controller not only avoids the synchronization but also it suppresses any extra power absorption during the payback period. During this phase, the aggregate power of TCLs may exceed the nominal power consumption in order to let the devices achieve the average temperature. However, in doing so, it prolongs the time it takes for the devices to recover a steady state temperature; this method also prevents a rapid control of the power levels.

The possible applications for thermostatic loads are not limited only to the frequency response. In fact, the traditional temperature-threshold controller can be modified to allow the tracking particular power profiles; this way, TCLs can adjust their aggregate power absorption to enhance the generators dispatch by shifting consumption from one interval to an adjacent one (see Section 1.2). The real time alteration of the regular temperature deadband to track power profiles has been proposed by [41, 42, 33]. In general these works rely on the existence of a real time communication infrastructure; the implementation of such an infrastructure would be extremely costly considering the very large number of devices involved. Another implementation [34] has had to rely on several analytical and numerical approximations that ultimately complicate the control framework.

1.3.3 Impact of thermostatically controlled loads on power system scheduling

The third aspect to be considered is the impact that flexible energy/power consumption and the provision of various frequency services from thermostatic loads have on the system scheduling. Initial research has calculated the value for energy-reserve markets that include reserve services from the demand side. However, the limitations in the reserve supply did not reflect the physical constraints of any actual demand side technology (e.g. battery storage, electrical vehicles or thermostatic loads); hence, it has been often difficult to ascertain the feasibility of the solution obtained [43, 44, 45].

Recent works represent a turning point as they specifically consider thermostatic loads; in particular, Aunedi et al. [46] assess the value for system scheduling based on the TCLs' deployment of primary response. If the economic benefits highlighted in this work regard the improvement of the system security, other works [34, 47] quantify the economic value resulting from the adjustment of the TCLs' energy/power profile in response to variations in the electricity prices. Similar studies are performed in [48] where the TCLs are assumed to have variable power input and, therefore, they lose their peculiar *on-off* characteristic. The aforementioned works consider the ability to provide system services individually. Studies that quantify the benefits of multi-service allocation have considered only other technologies (e.g. storage) [26].

Due to the lack of satisfactory level of TCLs' controllability over different time scales [49], the reliability associated to the delivery of response services from TCLs (e.g. compared to generators) is still a binding issue for active participation of TCLs in power system scheduling. As a consequence, it rises the need for an appropriate platform for TCLs that enables a reliable optimal allocation of a variety of services (i.e. frequency services and energy arbitrage) to be delivered simultaneously.

1.3.4 Current experiences and future practices with Demand Side Response: the TSOs' perspective

This section aims to offer the state of art review regarding the current demand side response practices adopted by several TSOs.

National Grid (UK): This TSO dedicates two balancing services for demand side actors. The demand customers that provide frequency control by demand management (FCDM) [50] *"are prepared for their demands to be interrupted for a 30 minute duration. FCDM is required to manage large deviations in frequency."* The second service is called Demand Side Balancing Reserve (DSBR) [51] and "*is looking to sign up large energy users who have the flexibility to reduce their electricity use when demand is at its highest between 16:00hrs and 20:00hrs on winter weekdays"*.

PJM (USA): the section "Demand Response" from the TSO' web page [52] attests that *"Demand response resources must install infrastructure to allow them to curtail their consumption of electricity within 10 minutes; they will be requested to curtail only when system conditions require the 10-minute response".*

NYISO (USA): the formal description of the NYSO emergency demand response program (EDRP) is on the web page [53]. This program *"is deployed in energy shortage situations to maintain the reliability of the bulk power grid. The program is designed to reduce power usage through shutting down of businesses and large power users."*

Hydro Quebec (Canada): The Demand-Side concept for the Canadian TSO is more focused on the energy efficiency [54].

None of the TSOs considered actually implements advanced frequency control schemes for DSR, as there is no possibility to perform a fully distributed load shedding to shut down only those loads that are relatively insensitive to power reductions (e.g. TCLs). Most of the services requested by TSOs regard day-ahead demand response programs. Therefore, as current practice the power system security is still largely ensured by generators or, in extreme cases, by generic load shedding procedures, which consist in expensive and indiscriminate disconnections of several substations across the network.

The high demand response potential of thermostatic loads has been recognized by the European TSO (ENTSO-E). In the recent "Network Code on Demand Connection" [55], two frequency-controlled (autonomous) pathways for the specific provision of frequency services by thermostatic loads are designed (See Article 23 and Article 24 of the code). For the first time, the role of temperature controlled devices in the provision of frequency response is highlighted. This code may serve as a basis for future deeper integration of TCLs within national TSOs' ancillary services schemes.

1.4 Thesis scope and original contributions

In this section the scope of the thesis and the original contributions to the existing literature are summarised.

1.4.1 Scope of the work

The research area of this thesis is the operation of future low-carbon power system using sources of flexibility and control alternative to conventional generators. Various options contribute

to the concept of Smart Grid; in particular, this thesis investigates the potentials of demand side response support by means of thermostatically controlled loads.

The scope of the work is to quantify the impact on the commitment and dispatch of a future low-carbon (i.e. with reduced inertia) system if a large population of thermostatically controlled loads is optimally controlled to simultaneously provide frequency response services and other forms of system services (e.g. energy arbitrage, congestion relief). The objective is to enhance the power system operation and, at the same time, to support the strict compliance with network security standards. In addition, the system support granted by TCLs' flexibility has, as further objective, the increased ability for the network to include RES in order to meet strict national and international emissions' targets. Moreover, the reliability associated to the supply of the various system services has to be strictly considered as, if not properly ensured, it would represent one of the most binding barriers for utilization of TCL's flexibility. This point suggests a further consideration; scheduling TCLs' provision of frequency services has to consider accurately the effect of the temperature/energy recovery. The system operator has to be aware of the TCLs´ recovery pattern and schedule appropriate reserve units to avoid possible frequency drops during the payback period. This important backup comes with a cost that has to be included in the calculation of the TCLs' value. Finally, the actions performed by demand side actors must not significantly affect the quality of the service of the primary function of the appliances i.e. maintaining the temperature of each device around a set point value.

The achievement of this objective requires the transition through two intermediate steps before being able to quantify the system level value with TCLs' control. First, it is required to establish relevant device level model. For instance, considering a refrigeration unit, splitting the dynamics of the food temperature and the cabinet temperature is useful to capture the fact that the temperature of the food changes much slower than that of the cabinet. This could potentially increase the time during which the devices can be switched off without actually affecting the food safety. Moreover, apart from modelling TCLs that belong to new sectors (e.g. heating systems, water heaters), it is important to recognize the existing differences between appliances within the same sector; the dynamics of a domestic refrigerator and its associated performance for demand side support may be quite different, even compared to those of a domestic freezer.

In addition, the intrinsic flexibility associated to a large number of TCLs would be handled more easily by means of a model that describes the dynamics of TCLs directly as a cluster rather

than considering the appliances individually. The second step deals with the need for a satisfactory control strategy that is able to accurately control the aggregate TCLs consumption over different time scales. A unique control architecture should govern both the short and long term time scales; TCLs would be therefore able to react to events within a narrow time window (seconds) and conduct a precise recovery phase to the regular operation. It would be even more beneficial if the same controller allows other DSR applications (e.g. energy arbitrage). Finally, the controller should guarantee at all times that the actual provision of any demand side service does not violate strict per-appliance temperature constraints.

1.4.2 Original contributions

In this subsection the original contributions of the thesis are summarizes. In Chapter 2 two new concepts have been introduced: “dynamic effective inertia” and “average effective inertia”. These intuitive concepts allow to evaluate the impact of the inertial support granted by controlling thermostatic loads, even though the conventional definition of an inertial constant is not pertinent to these appliances as they don’t store any kinetic energy. In addition, an increased amount of “effective system inertia” is revealed; this results in higher frequency nadir. Considering the concept of average effective inertia, it is possible to demonstrate the contribution of demand response compared to other solutions (e.g. conventional synchronous generators).

In Chapter 3 a fourth order dynamic model is proposed to describe the temperatures’ dynamics of a domestic combined fridge-freezer. This representation improves the dynamic models provided in literature that regard the same appliances. This model recognizes internal heat flows i.e. between the internal wall of the compartment and the contest stored in the same compartment. This way, the dynamics of the food are proved to be much slower than those related to the internal wall of the compartment (the state variable that controls the thermostat). The second order models for single freezers or fridges are derived from this general model; the use of second order models is due to the presence of a single compartment for these devices.

In this chapter simple but effective first order dynamic models of a number of commercial refrigeration units are derived; in particular, the study includes bottle coolers, commercial refrigerators, commercial freezers, and two different multideck refrigerators; such models could not be found in public. Furthermore, higher-order models for domestic refrigerators have been translated into first order models; since these models are sufficiently accurate in the nominal range

under consideration, they do not introduce significant errors compared to higher-order solutions. The analysis of the collective response of thermal devices prompts some key observations that have been used to define system-level performance criteria for initial controllers. Therefore, a hybrid controller has been developed addressing the shortcomings observed with other solutions [29, 31, 28]. The design of this novel controller complies with the defined performance criteria and it is tailored for frequency control applications.

Chapter 4 deals with an advanced control strategy for thermostatically controlled loads. This chapter builds upon the collaborative research results presented by Tindemans, Trovato and Strbac in [56]. Sections 4.2 - 4.7 consider the original results, whereas Section 4.8 presents new and additional results . My contribution to that work regards (a) the definitions of the system-level and device-level requirements, (b) the descriptions of device models, (c) the definition of the control challenges and (d) the active discussion on high-level framework of the control architecture. The core of the control strategy is the ability to track arbitrary power profiles (within limits) using independent actions of a heterogeneous cluster of TCLs. This result enables a number of demand response applications (frequency response and energy arbitrage) avoiding the need for an expensive and unreliable communication infrastructure. Moreover, it is demonstrated that this controller strictly respects device-level temperature constraints. The proposed methodology is able to simultaneously provide four features for advanced demand side response applications that could not be granted by alternative solutions in literature. I also evaluated the impact of two different implementations of the advanced controller on the network frequency control (see Section 4.8).

The original contribution of Chapter 5 is the modelling of a *leaky storage unit*, which describes the aggregate flexibility of a heterogeneous population of TCLs as a single storage unit with a leakage/evaporation term. This intuitive representation has the advantage of facilitating the embedding in an optimal service allocation framework. The leaky storage model incorporates the physical thermal models of heterogeneous TCLs and their operational temperature limits. The model is exact for large populations of TCLs that can be described by constant first order models and – in combination with a suitable controller – it guarantees that scheduled responses can be delivered by the least flexible appliance. Dr. Simon Tindemans has contributed to achieving some of the results presented in this chapter; his contribution is limited to the to the translation of the control strategy to the leaky storage model.

The leaky storage model introduced in Chapter 5 offers a novel representation of the TCLs that is very convenient for the system-level studies carried out in this Chapter 6. The contribution of this chapter is the design of demand side response platform that is able to guarantee the simultaneous deliverability of energy arbitrage and frequency response services. Moreover the flexible TCLs consumption enables the provision of transmission constraint management service. The value for TCLs' flexibility is analysed in the context of a SCED model that optimizes the generators' output and TCL energy consumption in a two bus-bar system, subject to frequency response requirements and network power flow requirements. Hence, the contribution of this chapter represents the first attempt to quantify the system level economic value of the intrinsic flexibility of TCLs.

Chapter 7 provides further original contributions. The DSRM introduced in Chapter 6 is enhanced as it includes the dynamics of the load recovery pattern and calculates the associated post fault energy levels, previously ignored. After the provision of secondary response, the recovery causes the TCLs to absorb an extra amount of power, compared to the regular power consumption scheduled for that time interval. In this new version the optimization problem does consider the effect of the TCL energy recovery on the system operation cost. Moreover, the DSRM is based on novel characteristics, or rather, the premise that the energy deployed while providing frequency services at the generic time interval has to be fully paid back by the end of the same interval. This way TCLs would always be able to guarantee the scheduled energy profile and the ability to provide the scheduled response services at all intervals regardless of the occurrence of failures at previous time steps. This characteristic makes the TCLs' level of service comparable to generators and makes the demand side supply of security services highly reliable. The proposed DSRM is integrated within an advanced SUC routine in order to quantify the value for the system if TCLs are controlled to provide a number of system services. The integration of the DSRM allows for system operational cost savings and increased ability for the network to absorb RES. The initial version of scheduling model refers to a joint paper with Fei Teng [57]. I personally contributed to [57] with the mathematical formulation of novel inertia-dependent constraints on system frequency. These will affect the commitment and dispatch decisions so that the dynamic evolution of post-fault frequency will satisfy security requirements associated with the maximum RoCoF, the frequency nadir and the intermediate steady state frequency limit, with respect to the GB security standards. Chapter 7 also proposes a novel and insightful discussion regarding the impact of solving the SUC with rolling

planning on the respect of a TCL average energy constraint; in particular this constraint refers to the safeguard of the primary function of the appliances. Three solutions are evaluated and discussed.

1.5 Thesis outline

The remainder of the thesis is organizes as follows:

Chapter 2 describes the potential of demand side contribution to enhance the system inertial response in a low carbon power system with a large share of renewable energy sources. Two interesting concepts are introduced: the “dynamic effective inertia” and “average effective inertia”. The distributed support by TCLs is translated into an increase of the effective system inertia that facilitates the integration of large shares of RES, otherwise curtailed. The contribution of demand response is compared with another solution (i.e. inertial response of conventional generators). Finally at the end of the chapter, it is calculated the minimum demand side penetration necessary to operate the system as it is virtually operated only by synchronous generators.

Chapter 3 is divided into two parts. In the first one the dynamic thermal models of domestic and commercial refrigeration units are introduced. High order models are computed for domestic cold devices whereas first order model are derived for commercial appliances. The focus of the second part is on the control of the aggregate TCLs’ power consumption; hence, two controllers from literature are investigated, pointing out undesirable side effects at device or system level. A hybrid controller is introduces to address the observed shortcomings. Quantitative comparisons are made by means of computer simulations.

Chapter 4 deals with an advanced controller for flexible demand response with thermostatic loads. The first sections (4.2-4.7) make a summary of the high-level properties of the control framework; four main features for demand side response applications with TCLs are provided by the presented controller. The high-level properties of such control architecture allow for the design of two implementations that enhance the system frequency control; with the first option TCLs target a pre-programmed power curve in response to frequency events. The second instead makes the TCLs aggregate consumption proportional to the system frequency evolution and/or to its rate of change. This chapter also quantifies the impact of both these two implementations on the system frequency control under a future low-carbon scenario.

Chapter 5 investigates the optimal simultaneous allocation of a number of system services that maximizes the revenues of a cluster of TCLs. The initial focus is on the development of an envelope model that, starting from physical device model parameters and the associated constraints on the quality of the service, evaluates the aggregate flexibility of a heterogeneous population of TCLs. This way, it is straightforward to illustrate the different capabilities of various classes of cold devices. Afterwards, an equivalent representation of a cluster of devices as a leaky storage unit is introduced. It is embedded in a linear optimization tool to calculate the optimal multi-service portfolio for a large population of TCLs.

Chapter 6 changes the point of view of the analysis carried out so far; the technical achievements are used to assess the value for the power system associated with the TCLs' ability to perform energy arbitrage, to provide frequency response services and transmission constraint management services. The TCLs' ability to provide the first two types of services has been introduced in the previous chapters, whereas economic aspects associated to the last service are discussed in this chapter. In this chapter a multi-services demand response model (based on the leaky storage unit) is integrated in a SCED that minimizes the system operational cost of a two bus-bar system, subject to frequency response and transmission constraints. Further sensitivity studies quantify the impact of different penetration levels of controllable loads and transmission flow constraints.

Chapter 7 introduces an upgrade of the DSRM for the simultaneous provision of several system services already investigated in previous chapters. The model now accurately includes the effect of the loads' energy recovery and improves the reliability associated to the provision of response services from TCLs. The cost of supplying extra power during the payback phase is taken into account and included in the final calculation of the value for TCLs' system support. The DSRM is embedded in an advanced SUC routine particularly suitable for low carbon power systems. Several case studies are run to quantify the value of optimally rescheduling the TCL energy consumption in order to reduce the system operation costs and reduce the overall wind curtailment. Moreover we focus on the effect on system cost savings and wind integration if the response services (primary and secondary) response are provided individually or simultaneously; the comparison is also enlarged considering different generators response capacities. A further study analyses the sensitivity of cost savings and wind curtailment to the primary response delivery time; in addition, interesting results show the sensitivity to the TCLs recovery pattern, considering the generation mix

available for reserve services. Furthermore, we drastically increase the penetration of TCLs in the GB system and we obtain promising outcomes towards the abatement of greenhouse gas reduction. Finally, we discuss the impact of different formulation of an average energy constraint on the daily and annual operation of TCLs; the application of rolling planning in the SUC solution introduces important issues to consider.

Chapter 8 summarizes the conclusions of the thesis and discusses the extensions of this work.

Chapter 2: Initial Assessment of Power System Support Using Demand Side Response:

Abstract

This chapter describes the potential demand side contribution to enhance the inertial response of a low carbon power system displaying a large share of renewable energy sources. As a matter of fact, these non-emitting technologies, although capable of helping the network to meet impelling environmental targets, they do not currently grant inertial response, thus likely pushing the system operation in a dangerous territory. This chapter assesses the potential benefits for system security deriving from the procurement of *'effective inertia'* by controlled domestic thermostatic loads. This inertial support mechanism would facilitate the integration of available wind power even in critical scenarios characterised by low demand and high wind penetration.

2.1 Introduction

In the previous chapter we introduced the general context that motivates the research problem to develop in this thesis. In particular, we discussed about the aim of integrating large shares of renewable energy sources (essentially wind generation in UK) to reduce greenhouse gas emissions. Hence this chapter deals with the effect on frequency dynamics (RoCoF and nadir value) in a network characterised by reduce level of inertia due to RES online.

The majority of the system demand has been historically covered by conventional synchronous generators. Moreover, it is vital to a proper network operation to guarantee that the generating units' output equals the electric loads' consumption at all times; in fact, this situation implies the equilibrium between the mechanical torque sustained by the generators' turbine and the counteracting electromagnetic torque imposed by the load. If that is the case, network frequency value matches its nominal value of 50 Hz [58]. However, this balance can be interrupted by major occurrences: on the one hand, small disturbances determined by recurring variations of the electric demand during the system normal operation; on the other hand a major event such as the outage of a generating unit. Proceeding through the chapter we will solely take into account the effect on system frequency due to the sudden loss of a large generator unit.

Conventional generating machines are sensitive to any fluctuation in the network frequency; therefore, in the unlucky event of a generation loss, the network will receive an injection of active power '*naturally*' supplied by the remaining synchronous units through the deployment of kinetic energy stored in their rotating masses. Hence, this natural phenomenon undoubtedly aids in limiting the system frequency rate of change. Large interconnected power systems have generally benefited from substantial system inertia due to large rotating masses associated with conventional units. As a matter of fact, so far, the level of system inertia and its effects on the RoCoF has only concerned the TSOs in charge of relatively small and isolated networks.

As previously mentioned, the generation portfolio that has historically characterised power systems around the globe is being drastically affected by the increasing share of RES, mainly from wind energy. Wind turbines, apart from the fixed speed induction generator type, avail themselves of high-power converters that essentially decouple the wind turbines from the AC power system [19]. Hence, this generation technology does not *naturally* contribute to the overall system inertial

response through mechanical coupling [18]. As the wind penetration increases, the system's inertia declines; the power imbalance after a sudden loss of a generator unit will cause larger fluctuations in the system frequency and a larger rate of change of frequency. Consequently, the risk of severe consequences for system security and reliability rises.

A focus on the GB context [11] confirms large impending transformations to the GB generation portfolio, including:

- Significant raise of renewable energy sources (24.1 GW of wind generation for year 2020)
- Larger maximum possible infeed generation loss (1.8 GW from 2014)

These two points define the constraints for keeping the system secure, in particular with respect to the regulation of system frequency. In particular, 1.8 GW represents the maximum capacity of the nuclear power plant that will be installed in the GB network. In effect, the former limit of 1.32 GW could act as a barrier to the timely connection of large generating units (as pointed out by EDF and accepted by OFGEM [59]). It is worth mentioning that the new limit has been adopted since April 2014 even if the 1.8 GW nuclear reactor (III generation) is still not in operation [60]. This choice reflects the possibility of a large generation failure in conjunction with a possible simultaneous wind generation drop. Furthermore, the second most onerous infeed loss for the GB system is the loss of the HVDC interconnector with France that is normally operated with a power flow from France to England).

Figure 1 illustrates the effect on frequency evolution following a sudden generator outage when the network is operated with three different levels of inertia. The current level of inertia (green solid curve) still insures a secure system frequency evolution. If the system suddenly loses the maximum-size generator, the frequency nadir brushes the minimum limit of 49.2 Hz stipulated by the GB SQSS prescribed by OFGEM and adopted by NG, the GB transmission system operator [20]. Because of reduced inertia, the decline in frequency is much quicker, dropping below the security limit (black solid). Once the decline has occurred, the frequency can be restored to safe values only through the procurement of a large amount of primary response granted from the automatic action of generators' governors. Ultimately, with the lowest value of inertia (red solid), the frequency drops in a few seconds to 48.8 Hz (dotted line), triggering the NG defence plan called LFDD scheme; under this program, large percentages of system load are automatically cut off to

avoid, in the worst case, a blackout [21]. Due to the activation of such an expensive defence plan, the network frequency can actually starts to recover to acceptable values. Note that the minimum allowed value of frequency has been set to 49.2 Hz in order to be sufficiently distant from the value that activates large system demand disconnections, or rather the LFDD scheme. In effect, the key to all of these levels settings is that the frequency drop, consequent to the largest generation loss, must not affect the customer quality of supply. In this context, even if frequency accidently falls below 49.2 Hz, there still is a 0.4 Hz of buffer before customers start to be curtailed. As Figure 1 testifies, in future, with reduced inertia, frequency may directly fall to 48.8 Hz after the failure of a generator, drastically decreasing the quality of the supply, as large demand disconnections would be more frequently activated.

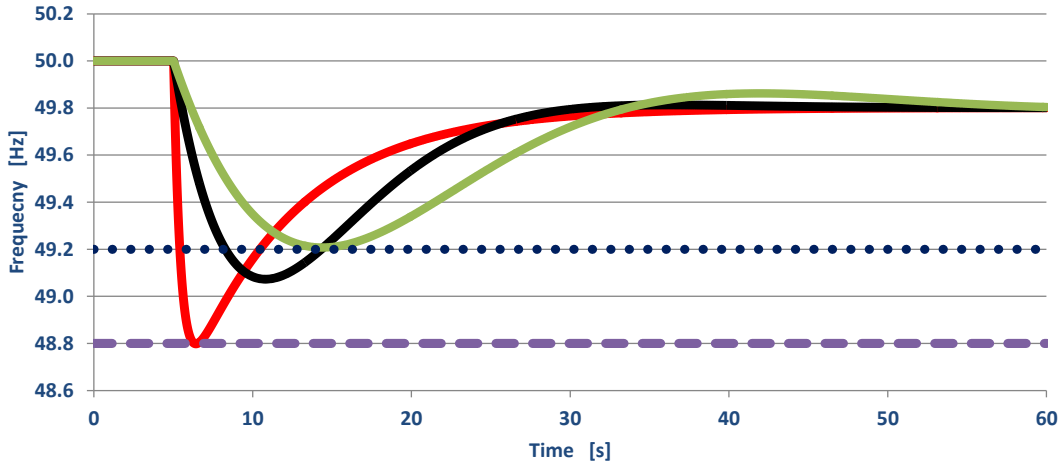


Figure 1 Effect of lack of inertia on frequency evolution: current inertia (green), reduced inertia (black), strongly reduced inertia (red); horizontal lines indicate the NG minimum frequency limit (dotted) and first stage of LFDD (dashed).

Since the system demand has been traditionally considered as uncontrollable, TSOs regularly operate the electricity networks so that system security is mainly assured by means of generators' control. In this vein, several works [27, 61, 62, 63] consider the provision of the so-called "synthetic" inertia by wind turbines, implemented through a supplementary loop in the turbine control scheme. Nevertheless, the ability of wind farms to bear the inertial support is still a limiting issue [64, 12]. Consequently, in sharp contrast with environmental purposes, the system operator often has no alternative but to limit wind production and commit emitting and high-cost conventional generators to supply the demand and provide additional response and reserve support. This way, the TSOs require conventional units to be operated part-loaded with a consequent significant lack of efficiency. Power system security and reliability remains therefore a priority and cannot be object of any allowance due to the integration of large portions of RES in the network operation; the need

for alternative solutions, aiming to reinforce the system inertial response in presence of decreased amount of natural inertia, is ever more impelling.

Demand side response represents a promising option to engage the aforementioned challenges. In this chapter, we suggest an illustrative case study attempting to show the potential beneficial effect of *smartly* controlling a large population of domestic TCLs so as to reinforce the inertial response of the power system under the GB 2020 Gone Green scenario. In the next chapter we will offer a more complete dissertation regarding the modelling and control of these devices, together with the distinctive features making the TCLs a suitable candidate for demand side response actions. For the moment, we focus on the potential impact of distributed TCLs' load shedding on the dynamics of the system inertial response. Specifically, we illustrate the concepts of *dynamic effective inertia* and *average effective inertia*. These concepts allow the quantification of the effect of inertial support procured by controlling thermostatic loads, whereas the conventional definition of an inertial constant is not pertinent to these appliances. In addition, we reveal the increased *effective system inertia*, ensued by the demand response, that results in higher frequency nadir. Later, we demonstrate the contribution of demand response compared to other solutions. Finally we calculate the minimum demand response penetration necessary to operate the system as virtually supplied solely by synchronous generators, fully compensating for the insufficiency of natural inertia from wind turbines.

This chapter is structured as follows. Section 2.2 focuses on the definition of system inertia and inertial response. Section 2.3 delineates the study context and the novel concepts we propose. Models used and scenario considered are outlined in section 2.4. Section 2.5 describes a simple dynamic demand controller design. In conclusion, section 2.6 summarizes the results of the study.

2.2 The inertial Response

The inertial response of a synchronous generator is described in [58] by Newton's second law of motion:

$$J\frac{d\Omega_m(t)}{dt} = T_m(t) - T_e(t) \tag{2.1}$$

Here J is the total moment of inertia of the turbine and the generator rotor [Kg m^2], Ω_m is the rotor shaft velocity [rad/s], T_m the mechanical torque produced by the turbine [Nm], T_e the

counteracting electromagnetic torque [Nm]. As the variation of Ω_m in respect to Ω_s (synchronous speed) is normally very small, that is $|\Omega_m - \Omega_s| \ll \Omega_s$, it follows that:

$$T_m\,\Omega_s \cong T_m\,\Omega_m = P_n \cdot p_m; \quad T_e\,\Omega_s \cong T_e\,\Omega_m = P_n \cdot p_e \tag{2.2}$$

P_n, is the nominal power of the machine [MVA], and p_m and p_e are respectively the relative mechanical and electrical power (both dimensionless). We define $f = \Omega/(2\pi \times 50Hz)$ where f is the dimensionless frequency. Using (2.2) this leads to:

$$\frac{df(t)}{dt} = \frac{\big(p_m(t) - p_e(t)\big)}{2H} = \frac{1}{2H}\Delta p_L(t) \tag{2.3}$$

$$H = \frac{1}{2}\frac{J\,\Omega_s^2}{P_n} \tag{2.4}$$

The inertia of a rotating mass resists changes in frequency through the deployment (or increase) of energy stored in the rotating shafts. The constant of inertia H, defined in (2.4), physically represents the time it takes for a generator to nullify the kinetic energy stored in its rotating mass, whilst still supplying the nominal power. It is expressed therefore in seconds and typical values for thermal generators are between 3 and 6 seconds [65]. Moving to a wider perspective, we consider a power system with N conventional generators. Thus, we derive the system equivalent of (2.3), where all the powers of each machine are referred to the system nominal power P_{sys} [MVA]. The system inertia results in a weighted average as follows:

$$H_{sys} = \frac{1}{P_{sys}}\sum_{i=1}^{N} H_i P_{ni} \quad = \frac{E_{sys}^{kin}}{P_{sys}} \tag{2.5}$$

Equation (2.5) highlights a useful interpretation of the system constant of inertia: E_{sys}^{kin} represents the energy reservoir due to the kinetic energy stored in the rotor shafts. At a nominal power P_{sys} this reservoir will last H_{sys} seconds. Operating the system with lower level of inertia reduces the size of the reservoir. Equation (2.3) with (2.5) illustrates the inertial response of a power system: in the first seconds after a generator outage, before the primary response fully acts, the slope of the frequency drop is determined by H_{sys} as Δp_L equals the amplitude of generation loss as a fraction of system load level.

2.3 Effective inertia for demand response devices

2.3.1 Overall considerations

The natural smoothing effect of inertia is beneficial for two aspects of network operation: reducing the frequency variations due to fluctuating imbalances between generation and load, and limiting the system frequency drop after a large disturbance. The former use is not easily addressed by responsive demand, due to the requirement for accurate and precise response to small changes in frequency. Instead, we consider the use of responsive demand to limit the drop in system frequency after a significant infeed loss.

Specifically we consider the contribution of domestic thermostatic loads such as fridges, freezers and fridge-freezers. In order to consider a generator or load as a real provider of inertia, a change in system frequency has to correspond directly to a change in its rotational speed and, thus, its kinetic energy. Clearly, the conventional definition of constant of inertia (2.4) cannot be applied to the responsive devices. Nevertheless, the capability of these devices to support system frequency can still be analysed in the inertial response time scale, i.e. before the full deployment of primary frequency response (~10 seconds).

2.3.2 Dynamic Inertia

Following a sudden infeed loss $\Delta p_L(t)$, the activation of demand response $p_D(t)$, means that the RoCoF is determined by the residual imbalance $\Delta p_L(t)$-$p_D(t)$. Because the permitted deviations from the nominal frequency of 50Hz are small, we continue to use the linearized equation (2.3) away from the nominal frequency. Hence, considering (2.5), we propose the definition of an instantaneous *dynamic effective system inertia* that takes into account the contribution of demand response on top of the actual system inertia H_{sys}.

$$H_{dyn}(t) = \frac{\frac{1}{2}\Delta p_L(t)}{\left(\frac{df(t)}{dt}\right)} = H_{sys}\frac{\Delta p_L(t)}{(\Delta p_L(t) - p_D(t))} \tag{2.6}$$

where $p_D(t)$ (positive) represents the contribution of demand response as power reduction (opposite sign to $\Delta p_L(t)$). If the control strategy, at device level, is able to elicit a simultaneous response from smart loads, $p_D(t)$ will resemble a step function.

Figure 2 illustrates the impact of demand response on the grid frequency. At $t = t_0 + \varepsilon$, the instant just after the generator loss, the dynamic demand contribution to system inertia is nil, so that the RoCoF follows from the mechanical inertia of the system. However, once triggered, demand response is capable of restricting the RoCoF, limiting the system frequency drop before the primary frequency services are fully deployed. In the relevant time between dynamic demand activation and primary response, the effect of this distributed smart load may be interpreted as an increased value of system inertia even though the number of synchronous generators has not changed.

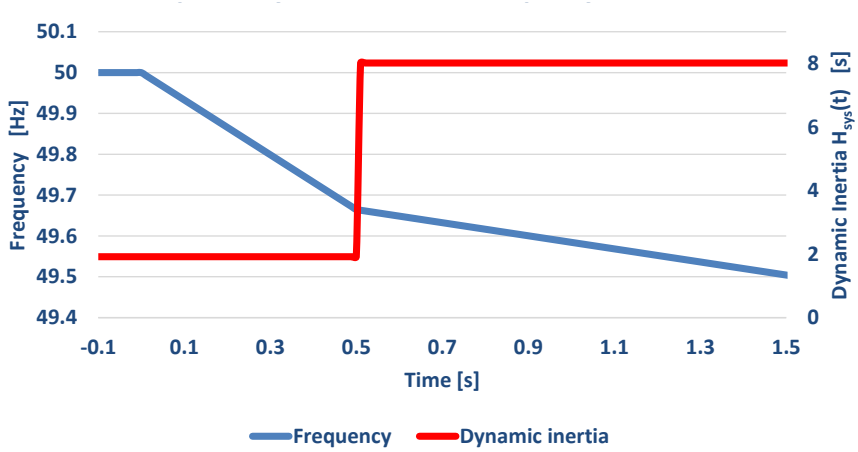


Figure 2 Effect of dynamic inertia on system frequency. The network scenario embodies 35 GW of demand, 20.1 GW of wind and 14.9 GW of conventional generators; considering (2.5), the resulting system inertia is H_{sys}=1.9s (wind farms do not provide inertial response). The simulated infeed loss is 1.8 GW whereas the responsive demand disconnects 1.37 GW (maximum aggregate power) at t=0.5s after the failure.

2.3.3 Average Inertia

Similarly, we introduce the concept of time-averaged effective inertia defined as:

$$\bar{H}(\Delta t) = \frac{\Delta p_L(t_0)}{2\left(\frac{\Delta f(t_0, t_0 + \Delta t)}{\Delta t}\right)} \tag{2.7}$$

Here, $\Delta f(t)$ represents the (linearised) frequency evolution of the network without contributions from frequency response services, and t_0 is the time at which the infeed loss occurs. Specifically, we will consider H_{10}=$\bar{H}(10s)$. This is the effective constant value that results in the same frequency value after 10s, for the same infeed loss. Figure 3 graphically illustrates the definition proposed.

We adopt a reference interval of 10 seconds; this represents the time for primary frequency response to reach its full magnitude [12]. H_{10} offers the system operator a single value for the effective system inertia that can be used to see whether the system is able to remain within frequency limits before primary response is ready to assist. Furthermore, the system operator, by fixing this value, may be able to set the dynamic requirements for the generators' primary response service, or rather, a secure system frequency time evolution.

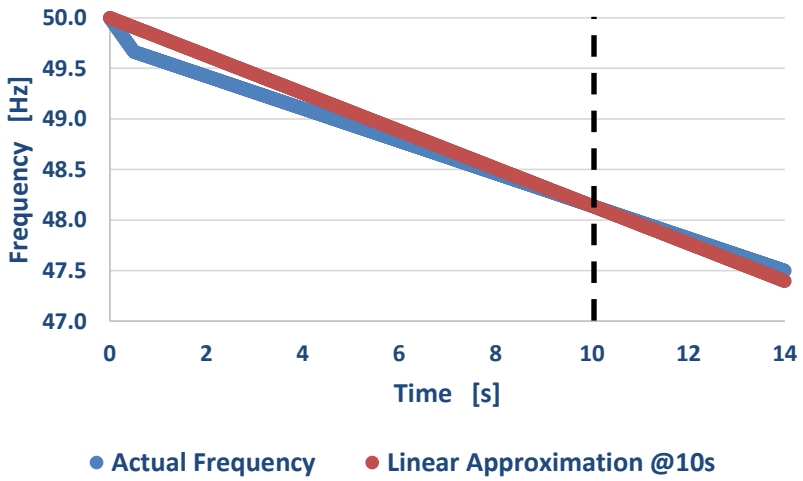


Figure 3 Comparison between the average inertia H_{10} with the dynamic evolution of H_{dyn} for 1.8 GW loss, 35 GW of demand and 20.1 GW of wind generation. Responsive demand disconnects 1.37 GW at t=0.5s.

2.4 Modelling and scenarios

This section outlines the scenarios simulated. We consider National Grid's Gone Green Scenario for the year 2020 [12]. The generation scenarios are characterized by a high, average and low wind penetration levels, respectively. These three cases illustrate how wind output can impact the frequency evolution and, more specifically, the first stages of the transient period. Each scenario embodies a certain amount of "must run" generation; this category mainly incorporates nuclear and wind generation. System demand varies from 20 to 55 GW, with 5 GW steps. Table 1 lists the must-run generation for each scenario.

Table 1 Must-run generation for 2020 GONE GREEN Scenario

Demand [GW]	"Must Run" Generation [GW]					
	Low Wind		Average Wind		High Wind	
	Wind	Synchronous	Wind	Synchronous	Wind	Synchronous
20	0	6.7	8	6.7	16.1	6.7
25	0	6.7	8	6.7	16.1	6.7
30	1.3	7.6	9.4	7.6	20.1	7.6
35	1.3	8.2	9.4	8.2	20.1	8.2
40	1.3	8.7	9.4	8.7	20.1	8.7
45	1.3	10.1	10.7	10.1	24.1	10.1
50	1.3	10.1	10.7	10.1	24.1	10.1
55	1.3	10.1	10.7	10.1	24.1	10.1

In few cases, if system demand is lower than the “must run” generation, additional demand (i.e. pumping) will be considered to clear the imbalance. On top of the “must run” generation, synchronous machines are assumed to feed the system. We assume the nominal characteristics of the thermal appliances (see Table 2), considering typical values for these devices [32, 28, 31]. The average power of a population of appliances is calculated as the product between the average nominal power of a single unit, the average duty cycle of the appliance and the total number of devices considered.

Table 2 Characteristics of responsive loads

Device	Average Nominal Power	Average Duty Cycle	Number of Devices	Average Power
Fridge	70 W	0.23	12.5 mil	0.20 GW
Freezer	100 W	0.30	13.0 mil	0.39 GW
Fridge-freezer	180 W	0.23	19.0 mil	0.78 GW
			Tot	**1.37 GW**

We extended the linear frequency response model in [66] to integrate the responsive demand and LFDD (Figure 4); it includes governor speed regulation (5% droop) and time constants representing the dynamic behaviour of the governor steam valve and the turbine (T_{z1}=4s; T_{p1}=40s). We used an average value of inertia (4.5s) for synchronous generators and no inertia for wind farms. The damping is introduced by E_c=1 (in units of P_n/50Hz).

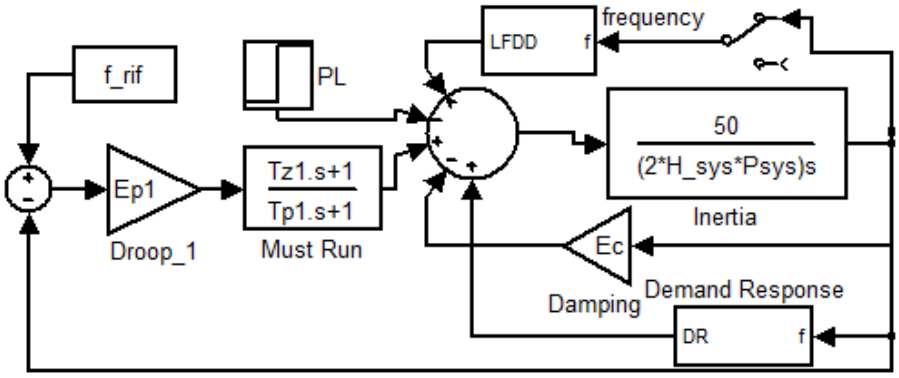


Figure 4 Linear system frequency response model

The variable P_L simulates a sudden loss of a single generating unit, set at 1.8 GW. If frequency drops below the 48.8 Hz threshold, LFDD will be activated. The parameters for this model are chosen to match the behaviour of the present GB power system and agree with the models and figures detailed by the National Grid [12]. A single busbar model is considered adequate to investigate the overall system inertia. The software used for the simulations is Matlab and its toolbox Simulink.

2.5 Control strategy for demand side support

Previous research [29, 31] offers some control strategies based on an overall provision of frequency services from demand response. However, the demand support is not sufficient in the inertial response time scale. In fact, the control action is more effective when frequency reaches the nadir (around 10 sec after the outage). In case of reduced inertia, frequency might have already achieved unacceptable values. The explicit use of the RoCoF in the controller would be helpful to enable a rapid dynamic demand response. The importance of making the control function for demand side response schemes sensitive to the RoCoF will be analysed in Chapter 3.

However, we now propose a simple decentralized controller design for the initial dynamic response of smart devices. The aim of the control action is restricted to the first seconds after the generator outage; Multi-tasking control algorithms are discussed in Chapter 3 and 4. Figure 5 offers the flow chart of the algorithm.

At the time that the failure occurs, each device in the on state identifies the frequency deviation and calculates its rate of change (RoCoF). This rate is used in conjunction with estimates for the conventional system inertia H_{sys} to evaluate the amplitude of the generation loss, Δp_L. The higher the sampling frequency the better the estimate of the power imbalance will be; moreover,

the estimates for inertia could be obtained from the system operator and updated periodically, for example through smart meters. Afterwards, the imbalance I is calculated as the ratio between the amplitude of the generation loss and the refrigeration average power p_{avg}; if this value is greater than 1, the appliance will immediately turn off.

If not, on the other side, a random number within an uniform distribution [0,1] is extracted and compared to I, as in Figure 5. Temperatures are not monitored here because our focus is exclusively on the initial frequency dynamics. On this time scale, the temperature remains essentially constant.

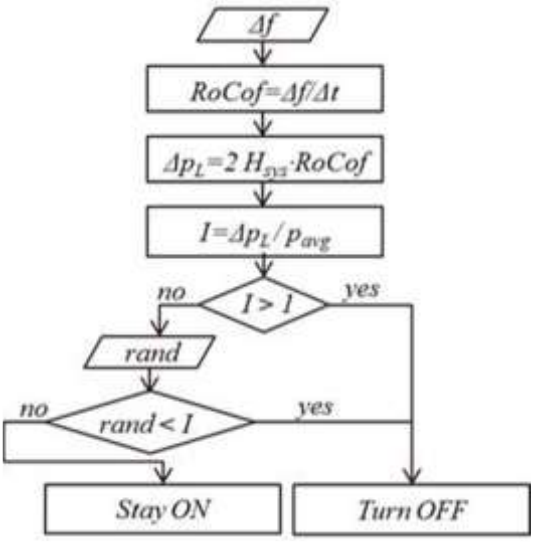


Figure 5 Flow chart of the proposed algorithm

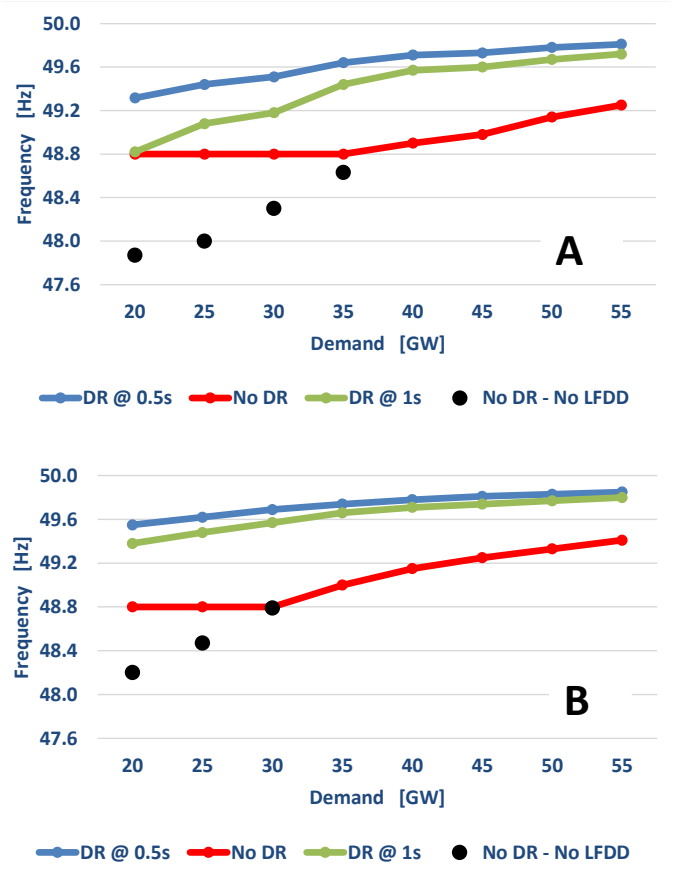
2.6 Results

In this section we evaluate the impact of demand contribution on the system inertial response. Table 3 classifies the values of the mechanical system inertia, H_{sys}, H_M, the maximum value of H_{dyn} evaluated as in (2.6) and H_{10} as in (2.7), considering an infeed loss of 1.8 GW and the ability to disconnect (with the control algorithm in Figure 5) all the responsive devices (1.37 GW) at 0.5s after the generator failure. The effective inertia of the system grows due to demand response support.

Table 3 Conventional inertia - Maximum dynamic inertia-Average inertia

Gen-Dem [GW]	Low Wind			Average Wind			High Wind		
	H_{Sys} [s]	H_{10} [s]	H_M [s]	H_{Sys} [s]	H_{10} [s]	H_M [s]	H_{Sys} [s]	H_{10} [s]	H_M [s]
20	4.5	16.5	18.8	2.9	10.1	12.2	1.3	4.87	5.5
25	4.5	16.5	18.8	3.1	11.4	12.8	1.6	6.1	6.7
30	4.3	15.6	18.0	3.1	11.3	12.9	1.5	5.5	6.2
35	4.3	15.8	18.1	3.3	12.1	13.8	1.9	7.0	8.0
40	4.4	15.9	18.2	3.4	12.6	14.4	2.2	8.2	9.4
45	4.4	15.9	18.3	3.4	12.5	14.4	2.1	7.7	8.7
50	4.4	15.9	18.3	3.5	12.9	14.8	2.3	8.5	9.8
55	4.4	15.9	18.4	3.6	13.1	15.2	2.5	9.3	10.6

Figure 6 highlights the observer frequency nadir as we vary the system demand and wind penetration. For each scenario we consider the full deployment of demand support after 0.5s and 1s or in absence of the service.



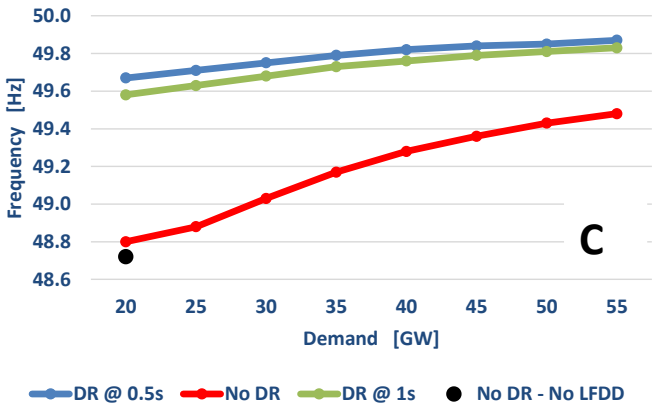


Figure 6 Frequency nadirs for (A) high wind, (B) average wind, (C) low wind scenarios. Results are shown for various demand response scenarios: full response after 0.5s (blue), full response after 1s (green) and no demand response (red). For this this case, data points at 48.8Hz are indicative of the LFDD activations. The black dots indicate the frequency nadir that would be reached without LFDD.

In the first case, for high wind penetration, the grid is not able to keep the frequency above 49.2Hz. The first stage of LFDD is triggered up to 35 GW of demand, and without it the frequency would have dropped to even lower values (black dots), potentially leading to a blackout. It is clear that the grid cannot integrate this level wind generation without curtailing wind. The situation drastically changes when demand response is activated. The minimum value of frequency never transgresses the security threshold, so wind generation can be integrated without further issues. This happens in all the cases with rapid support and most of the cases with "slower" support. With a slower support, the limit is passed only for high wind and very low demand. Nonetheless, the LFDD is never activated as frequency drops just above 49.2 Hz. The results improve as the amount of wind power generation decreases (Figure 6B and Figure 6C). The power system is secure even under low demand scenarios due to smart loads' contribution. With decreased wind availability, the need of a rapid supply of demand response becomes secondary.

Figure 7 demonstrates the contribution of demand response in the provision of effective system inertia compared to other solutions (e.g. conventional generators). Here we multiplied the mechanical and effective inertia by the system load to obtain a measure of (effective) energy. The effective energy $H_{10} \times P_{sys}$ is compared with the system's conventional kinetic energy KE_{conv} (sGW) obtained scheduling only synchronous generators for each scenario (H_{sys}=4.5s).

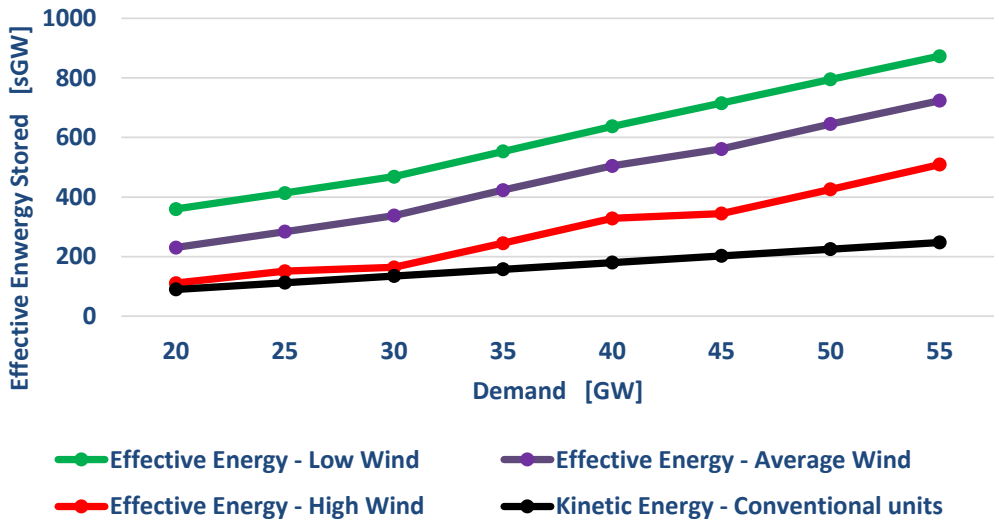


Figure 7 Kinetic energy stored in the system with or without demand response for different wind penetration compared. Parameters for demand response results: loss of 1.8 GW, demand response of 1.37 GW after 0.5s.

Alternative solutions for synthetic inertia usually result in numbers below the black curve, as, for instance, a controlled wind turbine offers a lower inertial contribution than synchronous units (per unit MW installed) [61]. All scenarios attest the decisive contribution of demand response.

Finally we evaluate the minimum penetration of smart loads, P_D^*, required to operate the network as "effectively" supplied only by synchronous machines whereas it actually integrates wind in each scenario. Considering Figure 2, we impose $H_{10}(P_D^*) = H_{conv}$ (i.e. 4.5s) and solve (2.7) for Δf_{10}^*; the new frequency $f_{(10s)}^* = f_0 - \Delta f_{10}^*$ is used in (2.8) to calculate H_M^*. As $f_{(0.5s)}$ is not varied, we find a new H_M^* as follows:

$$H_M^* = \frac{(\Delta p_L \cdot (10 - 0.5)s)}{2 \cdot (f_{(0.5)} - f_{(10s)}^*)} \tag{2.8}$$

hence, according to (2.6), we derive:

$$P_D^* = \Delta p_L \cdot \left(1 - \frac{H_{sys}}{H_M^*}\right) \cdot P_{sys} \tag{2.9}$$

Table 4 highlights the minimum percentage of the total refrigeration load (1.37GW) that should be responsive. Only in a few cases, a deep penetration of demand response is necessary. Hence, a gradual and more realistic penetration of responsive devices will not reduce the benefits highlighted.

Table 4 Minimum demand response penetration required.

		Generation – Demand [GW]							
		20	**25**	**30**	**35**	**40**	**45**	**50**	**55**
Min $P_{D\%}$	**LW**	0.0	0.0	6.0	5.7	2.6	2.9	2.9	3.1
	AW	48.1	42.1	42.7	36.4	33.2	33.5	30.3	27.4
	HW	96.7	87.4	91.5	79.3	70.0	73.2	67.1	60.9

2.7 Conclusions

This chapter offers an initial assessment of the benefits for the power system granted by DSR schemes; in particular, the positive demand side contribution by means of TCLs was evaluated over the short time scale of inertial response. A large amount of domestic thermostatic loads, controlled by a simple strategy, is able to quickly reduce its aggregate power consumption; such a distributed and quick load shedding improves the dynamics of the system inertial response and thus it allows the power system to integrate large shares of renewable energy sources, still preserving its security and reliability. This chapter proposes the 'effective dynamic inertia', a concept capable of extending the conventional inertia of synchronous machines and attaining greater values in the presence of demand response frequency support. Furthermore, the analysis advances the definition of 'average effective inertia'. Transmission system operators may utilize such a concept to set the future system requirements for primary response service, or rather, a secure system frequency evolution. A set of smart devices is proved to increment effective inertia and enhances the power system security during high wind scenarios, ensuring the most effective inertial support compared to other solutions.

2.7.1 Extensions and links with the next chapters

This introductory chapter serves as basis for further developments regarding the demand side contribution to the system operation and security. Next steps deal with the modelling of the thermostatic loads; as the quality of the service of these appliances is assessed through the ability to maintain the temperature of a compartment close to a set-point value, it becomes fundamental to ensure that the participation to system services does not impact the TCLs primary function. Moreover, the aggregate power consumption of such a large number of devices needs to be

accurately controlled not only on a short time scale (primary frequency response) but also throughout the whole frequency transient period. Furthermore, the design of satisfactory control strategies is not a trivial task due to the peculiar *on-off* characteristic of the TCLs operation. These aspects are investigated with more detail in Chapters 3 and 4.

Chapter 3: Modelling and Control of Thermostatically Controlled Loads

Abstract

This chapter deals with the modelling of thermostatically controlled loads such as refrigerators. It illustrates, by means of a fourth and second order dynamic models, the temperature evolution characterizing these appliances; first order model adaptations are also derived. These models are based on parameters determined analysing typical duty cycles and daily consumptions. The cited devices represent a particularly interesting subset of appliances for demand response support; the potential for short-term modulation of their aggregate power consumption, which allows the procurement of fast frequency regulation, is evident. The control of the aggregate consumption requires the design of control architecture; this chapter investigates the initial solutions suggested in the literature. These Initial controllers adjust the temperature deadband of each unit according to the frequency deviation; however, these frameworks are not immune from undesirable side effects at device or system level. After the analysis of the system-level performance criteria, we therefore introduce a hybrid controller that addresses the observed shortcomings. The effectiveness of this control strategy is illustrated through model system simulations. Two control schemes from the literature are used for quantitative comparisons.

3.1 Introduction

Demand side response includes several classes of responsive appliances e.g. common domestic and commercial appliances such as refrigerators, heating systems, dishwasher etc. The general idea underlying the abovementioned technologies is to quickly reduce the devices' aggregate power consumption in response to the outage of a generation unit and the resulting sudden frequency change. This fast and distributed load shedding contributes to re-establish the generation-demand balance, successfully arresting the decline of the system frequency within acceptable limits. This concept has been partially explored in the previous chapter.

Furthermore, during the regular power system operation, the devices' power consumption can be managed, shifting demand from peak to off-peak hours so as to re-shape the system demand profile (energy arbitrage). However, in this chapter we only focus on technical aspects related to the provision of frequency services after a significant frequency drop due to a sudden generator outage. Note that the ability for responsive demand to make energy arbitrage will be analysed and illustrated in Chapters 5.

In this chapter our attention goes to a particular class of responsive devices: the thermostatically controlled loads (TCLs). A detailed account of the standard operation of these appliances is provided in section 3.3, together with the mathematical models describing the main dynamics of these units. In particular, refrigeration units represent a substantial base load as they are connected to the grid at all times, making them suitable for providing frequency services around the clock. Moreover, these devices, if truly controlled, would be ready to participate to the frequency support as they are already installed across households and commercial sites. Although the contribution from heat pumps and generic electric heating systems would be significant, nowadays only 5% of the GB households make use of electricity as source of heating [30].

Another factor that complicates the modelling of heating systems deals with the wide differences in the households' construction; in particular, the level of insulation, which largely affects the temperature dynamics, could be significantly different varying the house typology. Finally, these appliances are not connected to the network at all times; a time dependent consumption would represent a further complication. Hence, in this chapter and in the remainder of this work we only focus on refrigeration units in both domestic and commercial sectors.

The primary function of these units is maintaining the devices' controlled temperature close to a set point value. This objective is achieved by varying the power consumption between two states; in the *on* phase the appliance drains its maximum power and reduces the controlled temperature. On the other side, during the *off* interval, the power absorbed is negligible and the temperature therefore increases. The switching between the phases is regulated by means of a thermostat that operates using a temperature's deadband around the set point temperature; when one of the two temperature boundaries is reached, the thermostat commands to change the status of the engine. The possibility to access only two power levels represents a peculiarity of these appliances compared to other demand side response devices. This characteristic of TCLs operation requires particular care in the design of a control strategy.

The variety of possible control architectures for TCLs can be divided in two major categories; the first is a centralised control management; in this case the appliances are assumed to be connected to a communications network and are able to receive and execute commands that are generated by a central processing node [32]. However, unanticipated events such as the loss a generation unit are not explicitly addressed; moreover, the actual implementation of such a framework would require costly infrastructures that would, ultimately, limit a widespread implementation. In the remainder of this chapter we will focus on decentralized solutions for the control of TCLs ; the idea of *smartly* controlling thermostatic loads in a decentralised way is not new; in 1979, Schweppe proposed to use such 'energy type usage devices' for the provision of frequency services [39] to the electricity grid. By monitoring the grid frequency, the appliances can react to frequency deviations by reducing (due to low frequency) or increasing (due to high frequency) their power consumption.

The simple implementation is based on heuristic modifications of the regular dead bands controllers; in fact, considering common refrigeration units, a negative frequency deviation from the nominal value implies an increase in the upper and lower temperature thresholds. As a consequence of this some TCLs in the *off* state that were about to switch on (as they were achieving the upper bound temperature) delay the state jump; on the other hand some other TCLs in the *on* phase and approaching the lower bound, change status (*on* to *off*) earlier than expected according to the regular operation. Hence, in response to a frequency drop, an overall load reduction is performed from the responsive TCLs contributing to the restoration of the frequency balance. Later, when frequency recovers to the nominal value due to the provision of additional power from

generators, the temperature thresholds equal the nominal values and therefore the TCLs population would approach again the nominal power consumption.

The interest in these applications has increased in recent years. Several academic works and research project analysed the potential value of generalised demand response and in particular frequency support [67, 28]. This value has been recognised also by transmission system operators [68, 55, 69]. The use case proposed by [39] was analysed in detail by Short et al. [29] and by Aunedi et al. [28] for domestic refrigeration units on the GB network. In particular, these works investigate the potential for a decentralized controller for TCLs that makes the thermostat's temperature thresholds linear functions of the system frequency deviation from its nominal value. From the analysis of the results it emerges that the amount of frequency response required from conventional units can be drastically reduced.

Another approach, suitable for fully decentralised implementation, is represented by the stochastic controller designed by Angeli and Kountouriotis [31]. The probabilistic methodology aim to control the achievements of desired distributions from the probability densities involved. The algorithm disables the traditional hysteresis controllers by means of stochastic switching of the TCLs compressor's states (*on →off* and *off→ on)*. Both methods [28, 29] and [31] provide promising results, but each also has peculiar drawbacks. The detailed description of the control methodologies mentioned and a full explanation of the intrinsic drawback associated are discussed in section 3.4 of this chapter. The analysis of the collective response of thermal devices prompts some key observations that we use to define system-level performance criteria for responsive thermostatic loads. In addition, we design the methodology of a hybrid controller that extends the deterministic controllers [29, 28]. This controller addresses the shortcomings observed with the other solutions [29, 28, 31] and complies with the defined performance criteria. The effectiveness of this method is illustrated in section 3.5 by means of simulations in which we compare the proposed controller with the deterministic and stochastic controllers. The conclusions to this chapter are addressed in section 3.6.

3.2 Regular operation of refrigeration loads

In this section we explain the regular operation of a refrigeration cycle of a thermostatically controlled load [70]. Four main components are involved in the vapour-compression refrigerant

cycle, the most common refrigeration cycle : a compressor, a condenser, an expansion valve and an evaporator, as shown in Figure 8.

Once entered in the compressor as a vapour, the refrigerant is compressed to the condenser pressure. It then exits the compressor at a relatively high temperature and cools down progressively condensing as it flows through the coils of the condenser by ejecting heat to the surrounding environment. Afterwards it enters a capillary tube where it is subject to a severe drop in pressure and temperature due to the throttling effect. The low-temperature refrigerant subsequently makes its way in the evaporator, where it evaporates by absorbing heat from the refrigerated space. As the refrigerant leaves the evaporator to re-enter the compressor, the cycle is completed.

The exact alternation of the phases previously described is regulated by a thermostat. The thermostat adjusts the amount of power absorbed for heating/cooling in order to keep device's controlled temperature close to a set point value.

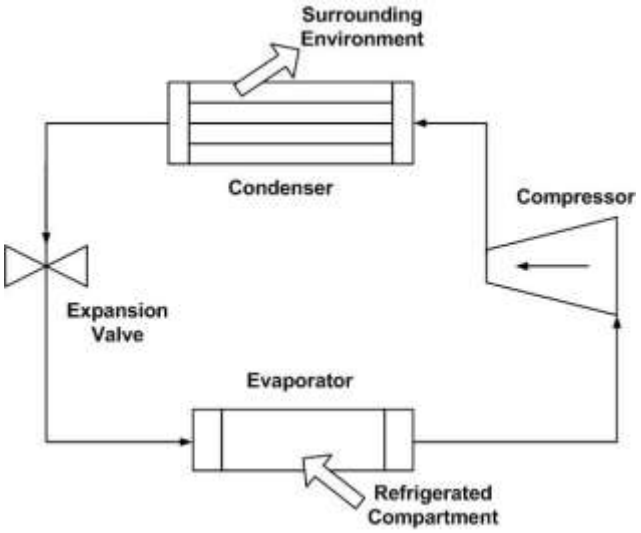


Figure 8 Basic components of a refrigeration system

In its most usual implementation, the thermostat makes use of a temperature deadband around the set point value. When the upper deadband threshold is overcome, the system switches to a cooling state by switching on the cooling mechanism (e.g. the compressor in Figure 8), and conversely, when the lower deadband threshold is overcome the system switches to a heating state. Hence, each individual TCL operates in two states, an *on* state when the appliance's engine is draining power, and an *off* state when the power consumption is insignificant. The cyclic repetition of both the phases for a domestic refrigerator is illustrated in Figure 9.

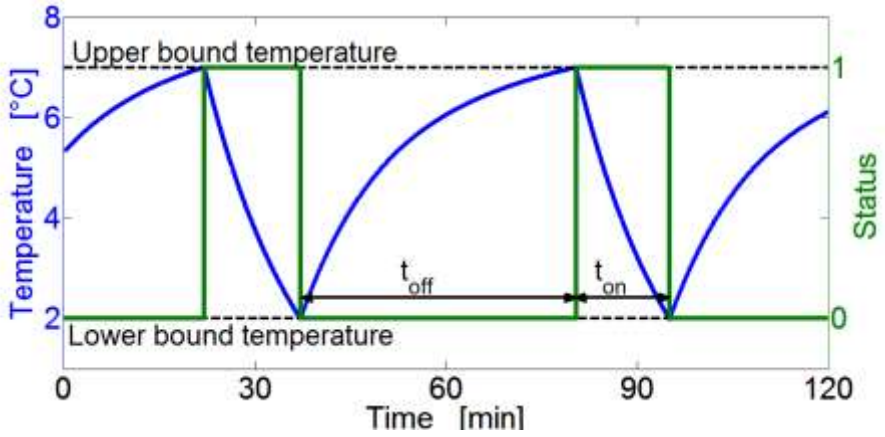


Figure 9 Regular operational cycle (green) and temperature evolution (blue) of a typical domestic refrigerator.

For this *cold* device, the active phase (Status=1) starts when the monitored temperature (green) reaches the upper bound temperature (dashed line); the power consumption and the cooling phase continues until the temperature reaches the lower bound temperature (dashed line). Afterwards the thermostat switches the compressor off (Status=0); the power consumption is now negligible and the temperature slowly increases towards the upper bound threshold , before that the thermostat changes the status again. The sum of the time durations of the *on* phase, t_{on}, and an *off* one, t_{off}, represents the period t.

$$t = t_{on} + t_{off} \qquad (3.1a) \qquad \pi = \frac{t_{on}}{t_{on} + t_{off}} \qquad (3.1b)$$

Another important parameter is the duty cycle π, defined as in equation (3.1b). In fact, the average consumption P [W] of a single device is obtained as in equation (3.2)

$$P = P_{on} \cdot \pi \qquad (3.2)$$

where P_{on} [W] is the compressor nominal power, roughly constant during the active phase. The higher the duty cycle is, the larger the consumption of the device will be; in fact, the duty cycle characterizes the performance of a TCL. Considering two refrigerators with the same P_{on} and same operating temperatures, the one with a lower level of insulation will require more time to let the temperature decrease, whereas the passive phase requires less time (again due to the lack of insulation). The duty cycle of this device is therefore higher than the one referring to a well-insulated appliance; the general performance is worse as the average consumption increases.

3.3 Dynamic models of refrigeration loads

This section deals with the formulation of the dynamic thermal model of refrigeration appliances that describe the time evolution of the state variables; in this case, the state variables are temperatures expressed in degrees Celsius. It is worth pointing out that a first order model would only describe the dynamics of the TCL controlled temperature i.e. the one monitored by the thermostat. Higher-level models are able to illustrate different temperature evolutions as they take into account internal heat flows, neglected in first order solutions.

3.3.1 Domestic refrigeration appliances

We now proceed to describe the temperature evolution of domestic cold appliances. Several works in literature model domestic refrigeration units by means of a first order ODE [31, 28, 32, 33, 34]. The unique state variable results to be a mixed description of the temperature of the internal wall of the unit's cabinet and the temperature of the food to chill. More specific formulations are offered in [29, 35]; these references consider a combined domestic fridge-freezer; a second order model therefore splits the dynamics of the fridge compartment from the freezer's dynamics. However, these solutions still do not separate the dynamics of the food temperature from those of the internal walls (where the measurement sensors are placed). Although the purpose of a refrigeration unit is to keep the food stored around a set-point temperature, the measured and controlled variable is the temperature of the internal wall of the unit.

We propose a general fourth order dynamic model to model a domestic combined fridge-freezer with a sufficient level of accuracy. Figure 10 shows the thermal model with the heat flows inside and outside a schematic fridge/freezer. H_1 represents the heat flow between the external ambient (the kitchen air) and the cabinet (fridge and freezer); the heat flow H_2 describes the interaction between the refrigerant and the cabinet (controlled by the motor). H_3 Illustrates the heat flow between the freezer cabinet and the freezer contents while H_4 between the freezer cabinet and the fridge cabinet. Finally, H_5 is about the heat transfer between the fridge cabinet and the fridge contents. The four state variables describe the dynamics of the cabinet temperature T_1 and the food temperature T_2, for the freezer compartment, and, again, the dynamics of the cabinet temperature T_3 and the food temperature T_4, for the fridge compartment.

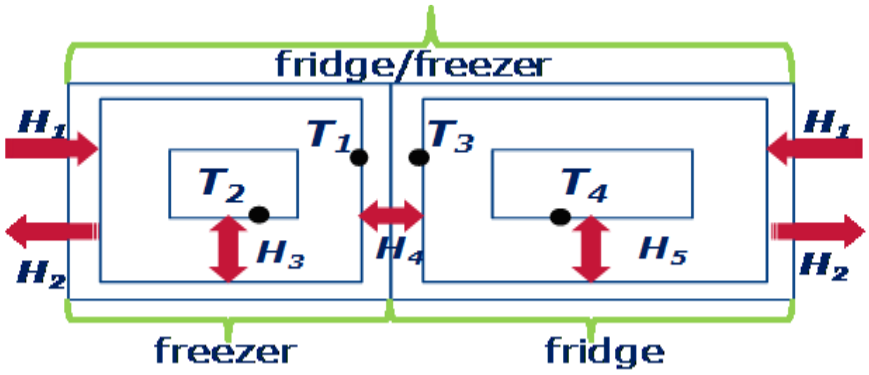


Figure 10 Schematic thermal model for a cold appliance with the heat flows involved.

The mathematical formulation of the thermal model is given by:

$$\begin{bmatrix}\dot{T}_1\\ \dot{T}_2\\ \dot{T}_3\\ \dot{T}_4\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & a_{13} & 0\\ a_{21} & a_{22} & 0 & 0\\ a_{31} & 0 & a_{33} & a_{34}\\ 0 & 0 & a_{43} & a_{44}\end{bmatrix} \cdot \begin{bmatrix}T_1\\ T_2\\ T_3\\ T_4\end{bmatrix} + \begin{bmatrix}b_1^{on/off}\\ 0\\ b_3\\ 0\end{bmatrix} \tag{3.3a}$$

It is worth pointing out that, in case of single freezer or fridge, the model simplifies to a second order model as there is only one compartment. The dynamics of the cabinet temperature T_1 and the dynamics of the food temperature T_2 (considering a single refrigeration unit or a single freezer unit) will evolve according to (3.3b).

$$\begin{bmatrix}\dot{T}_1\\ \dot{T}_2\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix} \cdot \begin{bmatrix}T_1\\ T_2\end{bmatrix} + \begin{bmatrix}b_1^{on/off}\\ 0\end{bmatrix} \tag{3.3b}$$

The parameters a_{xy} [s^{-1}] are defined as the ratio between the thermal conductance [W/K] and the thermal capacity [J/K], as defined in [70]. In addition, the parameters located on the main diagonal a_{xx} (see equation (3.4a)) are calculated considering the sum of the thermal conductances of all the y direct thermal paths with the variable x. The elements a_{xy} (3.4b) only consider the thermal conductance related to the path $x - y$. Both these elements are divided by the thermal capacity related to the variable x.

$$a_{xx} = -\frac{\sum_y K_{xy}}{C_x} \tag{3.4a}$$

$$a_{xy} = \frac{K_{xy}}{C_x} \tag{3.4b}$$

The definitions of $b_1^{on/off}$ and b_3 are:

$$b_1^{on} = \frac{-COP \cdot P_{on} + K_{1r}}{C_1} \tag{3.5a}$$

$$b_3 = \frac{K_{3r}}{C_3} \tag{3.5b}$$

Note that in b_1^{off} the power consumption P_{on} is nil and the subscript r reminds to the heat flow with the air in the room where the device is located. The values for these coefficients are obtained starting from reasonable initial values for the thermal conductance of the expanded polyurethane (the main insulating material) (K_p=0.03W/mK [70]), the internal convection coefficient (h_{ex}=40W/m^2 [70]) and the external convection coefficient (h_{in}=8W/m^2 [70]); regarding the thermal capacities, the specific heat of the cabinet is c_c=2kJ/kgK [70] and the water (c_w=4.2kJ/kgK [70]). We used the following dimensions and thickness of insulation: fridge: 1.8x0.6x0.5m, 4.5cm; freezer: 0.9x0.6x0.6m, 7cm; fridge/freezer: 0.6x0.7x0.6m, 8cm (freezer compartment) and 1.2x0.7x0.65m, 4.5 (fridge compartment). Furthermore we consider the weight of the contents: fridge 17kg; freezer: 20kg; fridge/freezer: 10kg for the fridge compartment and 4kg for the freezer compartment. Afterwards, parameters in (3.3) have been tuned to achieve a better match in accordance with desired average characteristics listed in Table 5 [71]. The ambient temperature is taken to be 20°C and it is assumed to be constant.

Table 5 Main characteristics of domestic thermostatic loads

Appliance	Power	COP	T_{max}	T_{min}	Duty Cycle
Fridge	70 W	2.5	7 °C	2 °C	0.240
Freezer	100 W	1.6	-15 °C	-21 °C	0.245
Fridge-Freezer	180 W	1.6	-14 °C	-21 °C	0.218

The numerical values of the model parameters are listed in Table 6.

Table 6 Nominal parameters of the dynamic model of domestic cold appliances

Fridge – second order model parameters [s^{-1}]				
a_{11}	**a_{12}**	**a_{21}=-a_{22}**	**b_{1on} [°Cs^{-1}]**	**b_{1off} [°Cs^{-1}]**
-7.05·10-4	5.56·10-4	1.04·10-4	-6.43·10-3	2.97·10-3
Freezer – second order model parameters [s-1]				
a_{11}	**a_{12}**	**a_{21}=-a_{22}**	**b_{1on} [°Cs^{-1}]**	**b_{1off} [°Cs^{-1}]**
-8.01·10-4	7.38·10-4	1.22·10-4	-8.37·10-3	1.30·10-3
Fridge/freezer – fourth order model parameters [s-1]				
a_{11}	**a_{12}**	**a_{13}**	**a_{21}**	**a_{22}**
-8.32·10-4	7.37·10-4	6.40·10-4	1.90·10-4	-1.90·10-4
a_{31}	**a_{33}**	**a_{34}**	**a_{43}**	**a_{44}**
6.41·10-5	-2.38·10-3	2.22·10-4	7.09·10-4	-7.09·10-4
b_{1on} [°Cs^{-1}]	**b_{1off} [°Cs^{-1}]**	**b_3 [°Cs^{-1}]**		
-10.60·10-3	6.10·10-4	1.90·10-3		

It is worth pointing out that the combined fridge-freezer model considers a unique thermostat, specifically referred to the freezer's temperature dynamics; as a consequence, these combined TCLs have only one compressor. By controlling the freezers' temperature dynamics, the temperatures of the fridge compartment are maintained within safe limits. Figure 11 shows the dynamics of the cabinet wall's temperature (solid blue) of a domestic refrigerator (second order model); note that this temperature is controlled by the thermostat in order to command the switching of the compressor status (see Figure 9). In Figure 11 the dynamics of the temperature of the food stored in this device (solid red) are also illustrated.

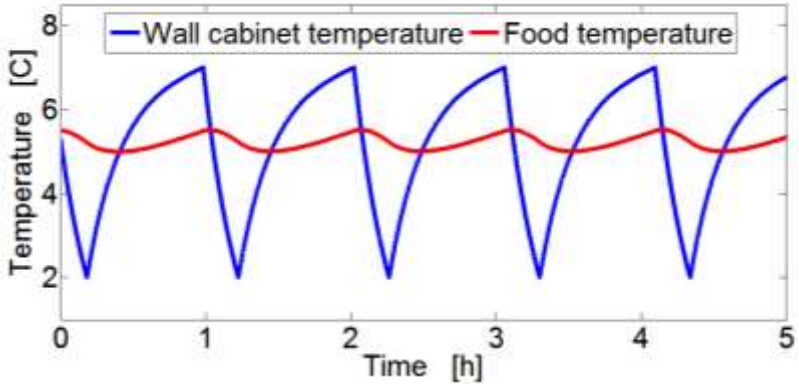


Figure 11 Temperature dynamics in a domestic refrigerator unit. The blue curve represents the temperature of the internal wall of the fridge cabinet and it is the controlled temperature. The red curve instead is the temperature of the food stored in the cabinet.

Splitting the dynamics of the food temperature and the cabinet temperature is useful to capture the fact that the temperature of the food changes much slower than that of the cabinet. This could potentially increase the time during which the devices can be switched off without actually affecting the food safety.

3.3.2 Commercial refrigeration appliances

The refrigeration appliances taken into account within the commercial sector are bottle coolers, commercial refrigerators, commercial freezers, and two different multideck refrigerators [72]. In particular bottle coolers are refrigeration units generally located in bars, pubs, coffee shops etc. The distinctive trait of these devices is the glass front door; regarding the multideck units instead, it is worth pointing out that these devices are open retail units. The main difference between the two multideck units is the size of these refrigerators. Examples of bottle coolers (left) and multidecks (right) are in Figure 12.

Figure 12 Examples of commercial bottle cooler (left) and multideck refrigerator (right).

The dynamic evolution of the commercial units' temperatures is described by first order ODEs. Although we recognize that higher order models would be more realistic and faithful, it is much more difficult to derive high order models for these appliances compared with domestic units. This complexity is mainly due to the architecture of these units, as the detailed effect of a glass or open wall (reduced insulation compared to the other regular walls) would be difficult to be modelled individually. Moreover, it is not easy to find detailed information in the devices' datasheets regarding the construction materials. However, first order models are still good approximations for initial analysis; the parameters for these models are computed considering typical duty cycles, power ratings of the compressors and daily consumption [73, 72]. Hence, letting $T(t)$ denote the temperature of a single appliance at time t, the evolution of the two operation states is described, with a first-order ordinary differential equation, as in [32]

$$\dot{T}(t) = -\frac{A}{m_c}\left[T(t) - \left(T_{off} - \frac{COP \cdot P_{on}}{A}\right)\right] \quad if\ ON \tag{3.6a}$$

$$\dot{T}(t) = -\frac{A}{m_c}\left(T(t) - T_{off}\right) \quad if\ OFF \tag{3.6b}$$

The temperature T_{off} [°C] denotes the ambient temperature, A is the thermal insulation [W/°C] and m_c is the thermal storage capacity [J/°C]; in addition, P_{on} [W] is the electrical consumption of the compressor and COP is the refrigerant's coefficient of performance. A more compact layout of model (3.6) is obtained by introducing τ, the thermal time constant [s] and T_{on}, the steady state temperature reached by a refrigerator which is always turned on.

$$\dot{T}(t) = -\frac{1}{\tau}[T(t) - T_{on}] \quad if\ ON; \quad \dot{T}(t) = -\frac{1}{\tau}\left[T(t) - T_{off}\right] \quad if\ OFF \tag{3.7}$$

$$\tau = \frac{m_c}{A}; \quad T_{on} = T_{off} - \frac{COP \cdot P_{on}}{A} \tag{3.8}$$

The temperature dynamics of the commercial refrigeration appliances can be significantly different when varying the device model. A commercial freezer is a device characterized by a deep level of insulation in contrast with a multideck that has an open surface (no insulation with the external ambient); although this structure facilitates the bars' customers, it penalizes the efficiency of these appliances. Compressors with large nominal ratings and high duty cycles are used in order to keep the food stored sufficiently refrigerated.

We now proceed to analytically derive the expressions of the duty cycle of an individual device making use of the first order model (3.7). The solution to the first-order ODE, depending on the compressor's status is given by:

$$T(t) = T_i e^{-\frac{t}{\tau}} + T_{on}\left(1 - e^{-\frac{t}{\tau}}\right) \;\; if\ ON; \quad T(t) = T_i e^{-\frac{t}{\tau}} + T_{off}\left(1 - e^{-\frac{t}{\tau}}\right) \quad if\ OFF \tag{3.9}$$

Let us now recall the definition of duty cycle expressed through equation (3.1b). Hence, we solve (3.9) for t_{on}, imposing $T_i = T_{max}$ and $T(t_{on}) = T_{min}$ (note that during the *on* phase the temperature decreases). Similar considerations are applied in order to calculate t_{off}, whereas, in this phase, $T_i = T_{min}$ and $T\left(t_{off}\right) = T_{max}$;

$$t_{on} = \log\left(\frac{T_{max} - T_{on}}{T_{min} - T_{on}}\right) \tag{3.10a}$$

$$t_{off} = \log\left(\frac{T_{min} - T_{off}}{T_{max} - T_{off}}\right) \tag{3.10b}$$

The duty cycle π is therefore expressed as:

$$\pi = \frac{\log\left(\frac{T_{max} - T_{on}}{T_{min} - T_{on}}\right)}{\log\left(\frac{(T_{max} - T_{on})\left(T_{min} - T_{off}\right)}{(T_{min} - T_{on})\left(T_{max} - T_{off}\right)}\right)} \tag{3.11}$$

The numerical values of the first order model (3.7) for the commercial refrigeration devices considered are listed in Table 7 [72].

Table 7 First order model parameters for commercial refrigeration units

Appliance	Power	T_{off}	T_{on}	T_{max}	T_{min}	τ	π
Bottle cooler	200 W	25°C	-1.03 °C	10 °C	4°C	5.54 h	0.70
Fridge	150 W	23°C	-28.22 °C	4 °C	1°C	15.18 h	0.40
Freezer	250 W	23°C	-84.47°C	-18°C	-22°C	19.16 h	0.40
Multideck 1	750 W	25°C	-5.42°C	6.5°C	1.2°C	0.59 h	0.70
Multideck 2	840 W	25 °C	-10.34 °C	6.5 °C	1.2 °C	0.96 h	0.60

3.3.3 First order thermal models for domestic refrigerators

Detailed thermal models for domestic appliances have been introduced in section 3.3.1. However, it is convenient to derive simple first order models that provide an overall but still effective approximation of the temperatures' dynamics of a refrigeration device. The possible lack of accuracy is counterbalanced by a large reduction of the computational burden and ease of manipulation in controllers. For the domestic units, default values for the temperature thresholds and room temperature are still the same as in section 3.3.1; the duty cycle and steady state cycle duration can be computed numerically using a complete fourth/second order model (3.3), and their values are used to fit T_{on} and τ for the first order model as expressed in (3.7). It is worth pointing out that the matching of a first order model to higher order dynamics results in unrealistic values for the asymptotic temperature T_{on} , but this approach provides the best fit in the nominal temperature range under consideration. The parameters for the reduced first order model are listed in Table 8.

Note that T_{on} and τ are bold typed to highlight that are numerically computed whereas all the other parameters do not differ from those introduced in Table 5.

Table 8 First order model parameters for domestic refrigeration units

Appliance	Power	T_{off}	T_{on}	T_{max}	T_{min}	τ	π
Fridge	70 W	20°C	**-44.1 °C**	7 °C	2°C	**2.03 h**	0.240
Freezer	100 W	20°C	**-134.6°C**	-15°C	-21°C	**5.13 h**	0.245
Fridge-Freezer	180 W	20 °C	**-151.8 °C**	-14°C	-21°C	**5.01 h**	0.218

3.4 Controlling the collective response of thermostatically controlled loads: performance criteria and initial controller designs

This section investigates the collective response of a large population of TCLs in response to a sudden frequency drop due to a generator outage. Hence we study the benefits and the drawbacks of using TCLs to support the system frequency control and consequent recovery to the steady state condition. For TCLs that are controlled using a temperature deadband, small fluctuations in

temperature are acceptable as long as the target temperature is approximately maintained over time. Because a large number of TCLs is connected to the grid at all times, tapping into the collective flexibility offered by these loads has large potential benefits. Figure 13 shows a schematic time evolution of the TCLs' aggregate power consumption after a frequency drop, starting from the steady state aggregate consumption level P_o.

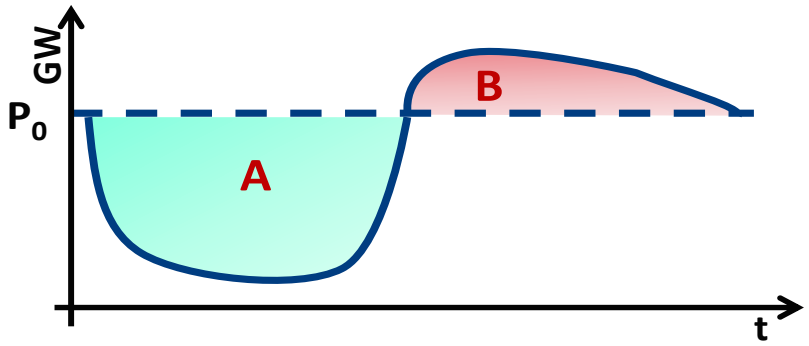


Figure 13 Time evolution of the aggregate power consumption of the smart devices. The curve starts from the steady state consumption level; after the power reduction and its consequent payback period, the aggregate power recovers the pre-fault level.

The area A represents the energy exploited by the power system and borrowed from the appliances; during this phase, a TCLs aggregate power reduction is enabled by frequency-sensitive controllers. These controllers adjust the temperature dead-bands of the individual appliances in accordance with the frequency drop so that, on average, the devices tend to spend more time in the *off* phase. This action helps the system to quickly balance the power deviation because of the lack of generation. Due to the deployment of the thermal energy stored in these appliances and the generators' governor response, the frequency deviation is contained and, after having achieved its negative minimum value, it starts to reduce. However, during this phase the thermal energy stored decreases and, thus, the temperatures of the TCLs increase. After the provision of the regulation support, the system frequency has recovered the steady state value of 50 Hz or lies in a small interval centred on that value. At this point, a frequency sensitive controller will re-establish the regular temperature thresholds, letting the TCLs recover the steady state duty cycle, and hence the power consumption. From this point onwards, the TCLs energy/power evolution is characterized by two distinct phenomena, the payback period and the synchronization. Although the appearance of these two issues may be simultaneous, they are not linked to each other and thus they have to be separately treated in the design of the control algorithm.

Payback Period: a steady state condition for a TCL appliance is defined not only by a steady state power consumption but also by a steady state average temperature (directly proportional to the thermal energy stored). At the time when the steady state power consumption is recovered (frequency back to 50 Hz and hence regular temperature thresholds applied), the temperature is still lower than the average value. In fact, area B represents the amount of energy that must be paid back in order to recover the devices' average temperature. This recovery phase is generally called *payback period*. During this phase the aggregate TCLs consumption may exceed the steady state level. The occurrence of this power boost can be the result of the inability of the TCLs' controller to manage the power consumption over this wider time scale or it can be included on purpose in the control strategy to speed up the temperature recovery (see Chapter 7). The power increase can be also entirely suppressed only at the expense of a very slow asymptotic recovery to steady state appliance temperatures (see Section 3.4.2 and 4.5.2 for the mathematical proof). For this reasons it is not possible to quantify a typical time duration for the payback period; moreover, a part from the controller design, this issue strongly depends also on the thermal inertia of the TCLs. Another crucial factor is the time during which the devices have been switched off (on average) in order to provide ancillary services; again, if TCLs are not fully controlled over long-term time scales, this time will vary depending on the actual on-line generation mix.

The extra power absorption is not necessarily a drawback for the system perspective as it is provided by reserve generators over a non-critical time scale (relatively low costs); it could be cost effective to increase the reserve power provision and diminish the frequency response required from generators (narrow time scale – high cost). This cost-benefit analysis will be investigated in detail in Chapter 7. As this chapter focused on the technical aspects of controlling large populations of TCLs, we limit our analysis to the controllers' characteristics and abilities. Finally, it is important to note that area B is smaller than area A: not all the borrowed energy needs to be paid back as a result of real energy savings in the form of reduced leakage losses.

Synchronization: the other issue that affects the TCLs provision of frequency control is the synchronization; this is actually a troublesome side effect caused by a large number of individual devices that, switched off to support the grid, do not automatically switch

back on when frequency recovers acceptable values. Instead, they are only switched on when their temperature reaches the maximum temperature threshold T_{max}. In addition, these appliances revert to an *on* state together at the same time with other appliances that were regularly *off* at the time of the frequency drop and therefore were not participating to the frequency support. All these appliances now *synchronize* their duty cycle and remain synchronized for a long time thus causing long-term power fluctuations with alternating phases of 'borrowing' and 'payback'. This issue can be solved introducing a sufficient level of randomness by means of stochastic switching during the TCL recover period. This way, the appliances do not achieve the *on* state together whereas they do it separately at different times. The disengagement strategies proposed within the controllers' design in [31] (see Section 3.4.2), in Section 3.4.3 and in Chapter 4 are all based on stochastic switching.

Previous research has demonstrated the potential of responsive domestic refrigerators to enhance the performance of the electricity grid. As anticipated in the introduction to this chapter, we specifically mention two control strategies for refrigeration TCLs, using different approaches: a deterministic control [29, 28] and a stochastic control [31]. Both methods work, but each of them also presents some drawbacks. In order to assess the benefits and the weak points of the mentioned approaches, we summarize the following three system-level performance criteria for responsive thermostatic loads:

- The ability to provide a sufficient reduction in power consumption at short notice.
- The capability to delay and control the payback of energy that the power system 'borrows' from the responsive demand.
- The ability to suppress synchronization of the duty cycles, which causes large-scale cyclic load patterns.

These three performance criteria are used as guidelines to design the proposed hybrid controller that extends the deterministic controller framework.

3.4.1 The deterministic controller

The idea of the deterministic TCLs controller comes straightforward from studying the operation of a single device: varying the maximum and the minimum temperature limits bring out

a reduction/increment of the average electrical consumption of the appliance. In particular, an increase in the temperature thresholds leads to a larger duty cycle and therefore to a larger power absorption (see equation (3.2)). The scheme in Figure 14A summarizes the design of the deterministic control.

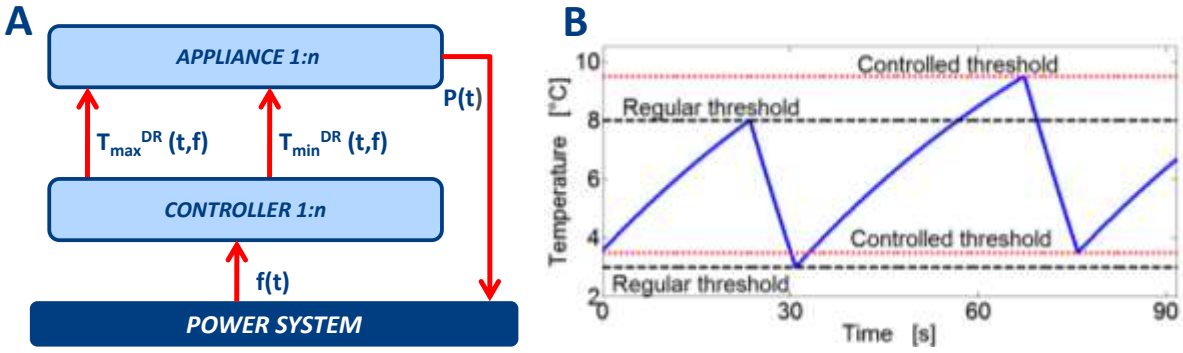


Figure 14 Schematic of the control algorithm of the deterministic strategy (A). Temperature evolution with modified temperature thresholds (B).

Let us consider a generation outage and the consequent frequency deviation from its nominal value; under this scenario, the decentralized controllers of each single appliance simultaneously identify the deviation and vary the thresholds of refrigerator's thermostat, $T_{max}^{DR}(t,f)$ and $T_{min}^{DR}(t,f)$, as linear function of the frequency deviation. In particular, the boundaries' update is given by:

$$T_{max}^{DR}(t,f) = T_{max} - K_f \cdot \Delta f(t) \qquad (3.12a) \qquad T_{min}^{DR}(t,f) = T_{min} - K_f \cdot \Delta f(t) \qquad (3.12b)$$

Where K_f is a positive gain, T_{max} and T_{min} are the regular temperature thresholds and $\Delta f(t)$ is the frequency deviation (negative value in case of a generation outage). Note that when frequency recovers the nominal value, and therefore $\Delta f(t) = 0$, or during the normal operation of the network, the default temperature boundaries are automatically applied. The prolonged *off* phase is shown in Figure 14B; the thermostat does not switch the compressor state when the regular threshold is reached as it waits until the temperature achieves the dynamically updated limit $T_{max}^{DR}(t,f)$. Although this methodology guarantees a quick response and a large power reduction in response to a frequency drop, it is strongly affected by the synchronization issue, as illustrated in the section 3.5.

3.4.2 The stochastic controller

We now proceed to recall the mathematical background supporting the stochastic controller developed in [31]; the underlying thermal model for TCLs is the same as for equation (3.7). We will now make a brief summary of the main results of this methodology in order to explain the most important properties and the objective of this algorithm. The devices are modelled as Markov-jump linear systems, where the Markov chain consists of only two states, *on* and *off*. This leads to the definition of two transition probability rates, λ_1 and λ_2. Note that in this framework the devices randomly switch between their operating states without any regular deterministic *on/off* switching operation. Furthermore, $\pi_{on}(t)$ and $\pi_{off}(t)$ denote the probability for a single unit of working in the *on* or *off* status, respectively. The sum of these two probabilities has to amount to one at all times. The time evolution of the aforementioned probabilities is defined by the following equations:

$$\begin{aligned} \dot{\pi_{on}}(t) &= -\lambda_1 \pi_{on}(t) + \lambda_2 \pi_{off}(t) \\ \dot{\pi_{off}}(t) &= -\lambda_2 \pi_{off}(t) + \lambda_1 \pi_{on}(t) \end{aligned} \tag{3.13}$$

At steady state,

$$\bar{\pi}_{on} = \frac{\lambda_2}{\lambda_1 + \lambda_2} \tag{3.14a}$$

$$\bar{\pi}_{off} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{3.14b}$$

It is worth pointing out that $\bar{\pi}_{on}$ also represents the average duty cycle of a single appliance. The transient probability distributions can be calculated by solving equations (3.13):

$$\begin{aligned} \pi_{on}(t) &= e^{-(\lambda_1+\lambda_2)t}\pi_{on}(0) + \left(1 - e^{-(\lambda_1+\lambda_2)t}\right)\bar{\pi}_{on} \\ \pi_{off}(t) &= e^{-(\lambda_1+\lambda_2)t}\pi_{off}(0) + \left(1 - e^{-(\lambda_1+\lambda_2)t}\right)\bar{\pi}_{off} \end{aligned} \tag{3.15}$$

Both the expressions above are monotone functions of time. If parameters $\lambda_1$and λ_2 change with a step function, the corresponding probabilities π_{on} and π_{off} will monotonically turn into new steady state values. The monotonicity of $\pi_{on}(t)$ is a first sign of the lack of power boost during the recovery phases, as the power consumption of a cluster of devices is proportional to the portion of them that are in the *on* state.

We now recall another important result derived in [31]. In fact, it is proved in this work that the expected value of the TCLs average temperature converges asymptotically to:

$$E[T(\infty)] = \bar{\pi}_{off} T_{off} + \bar{\pi}_{on} T_{on} = \frac{\lambda_1}{\lambda_1 + \lambda_2} T_{off} + \frac{\lambda_2}{\lambda_1 + \lambda_2} T_{on} \tag{3.16}$$

The analysis of the structure of equations (3.14a) and (3.16) suggests that the transition rates $\lambda_1$and λ_2 determine both the average temperature and the average power consumption of a collection of appliances. Hence, Angeli and Kountoutiotis propose a totally decentralized control of individual appliances that can be accomplished by choosing the transition rates $\lambda_1$and λ_2 as functions of the grid frequency deviation $\Delta f(t)$. This approach is summarized in Figure 15.

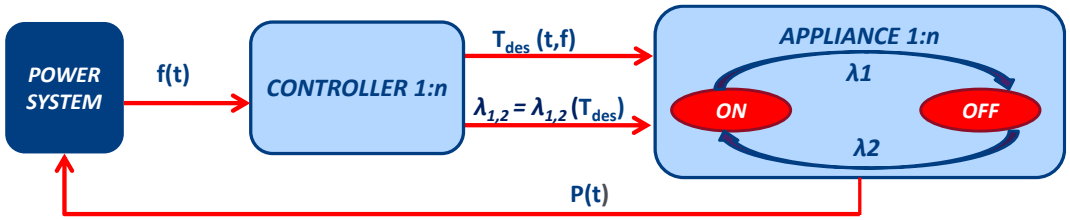


Figure 15 Control algorithm of the stochastic strategy

Each device´s controller identifies the frequency deviation and computes the desired average temperature T_{des} as linear functions of the frequency deviation.

$$T_{des}(t) = T_{nom} - K_s \cdot \Delta f(t) \tag{3.17}$$

Where K_s is a proportionality constant and T_{nom} is the nominal (target) value of the average temperature when frequency is at its nominal value of 50 Hz. Then, $\lambda_1$and λ_2 are settled as function of $T_{des}(t)$ (see [31] for the formulation). In the case of a negative frequency deviation, $\lambda_1$and λ_2 will be such that the number of TCLs jumping from the *on* to the *off* state grows.

The main objective of this control strategy is to avoid the synchronization of the individual appliances through a random switching and to suppress the long-term instability of simple deterministic solutions. The controller adapts the properties of a steady state distribution ($E[T(\infty)]$ and $\bar{\pi}_{on}$) and, thus, exercises a slow control over the temperature distribution and power consumption of a TCLs population. As a matter of fact, this design completely eliminates the payback phase, at the cost of an extended time for appliances to reacquire their steady state temperature. Moreover, this control architecture restricts the ability to implement rapid load changes (with respect to the first performance criteria previously listed). In fact, the positive results, such as the absence of synchronization, are achieved at the expense of limiting the speed and magnitude of the power reduction that is accessible to the system. Another drawback is the resulting on-off cycling frequency, which may entail excessive demands on the devices' engines.

3.4.3 The hybrid controller

The hybrid control strategy aims to achieve the three performance targets previously described and not fully guaranteed in [28, 29, 31]. During normal operation, TCLs operate within maximum and minimum threshold temperatures, respectively, T_{max} and T_{min}, resulting in alternating *on* and *off* phases. Hence, this method is based on the deterministic control, but it enhances the rapidity in the power reduction; furthermore it proposes a disengagement strategy to drastically reduce the synchronization by means of a random temperature T_{rnd} that commands random switching, alternative to the deterministic ones imposes by T_{max}^{DR} and T_{min}^{DR}. Figure 16A highlights the general framework of the algorithm while the left hand side (B) illustrates the anti-synchronization disengagement strategy. In general, the hybrid control is a decentralized approach as each appliance works individually.

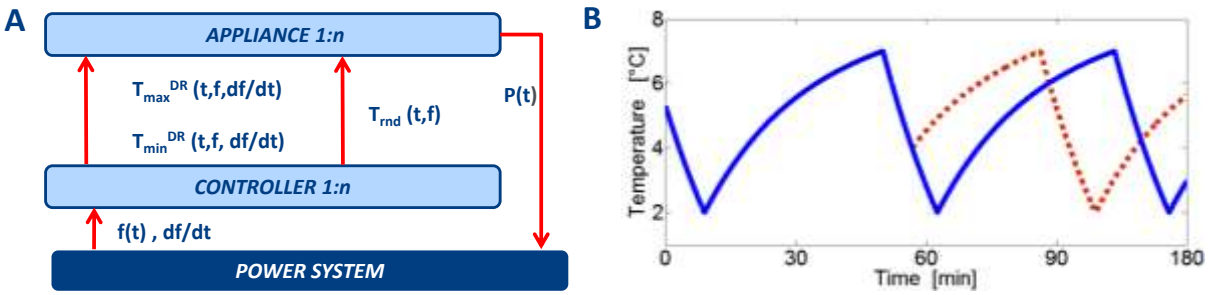


Figure 16 Control algorithm of the hybrid strategy (A). Illustration of the anti-synchronization disengagement strategy (B).

Each appliance switches on (off) when the controlled temperature reaches the modified lower (upper) limit T_{max}^{DR} (T_{min}^{DR}). The controller of each device identifies the frequency deviation $\Delta f(t)$, evaluates its rate of change, $d(\Delta f)/dt$, and updates the temperature limits according to:

$$T_{max}^{DR} = T_{max} - \left(K_f \cdot \Delta f(t) + K_r \cdot \frac{d(\Delta f)}{dt}\right)$$
$$T_{min}^{DR} = T_{min} - \left(K_f \cdot \Delta f(t) + K_r \cdot \frac{d(\Delta f)}{dt}\right) \quad (3.18)$$

By increasing the threshold temperatures when a frequency drop is observed, the devices collectively spend more time in the *off* state, reducing their aggregate power consumption. K_f is a positive gain that controls the overall sensitivity and K_r (positive quantity) controls the relative weight of the rate-of-change term. The values for these two parameters are empirically determined to produce a sufficiently large response whilst avoiding frequency oscillations. The introduction of

the rate of change in the update of temperature set points enables a faster response of the devices. In fact, Figure 17 shows that, with the hybrid strategy, the control action during the fast transient period is always maximum.

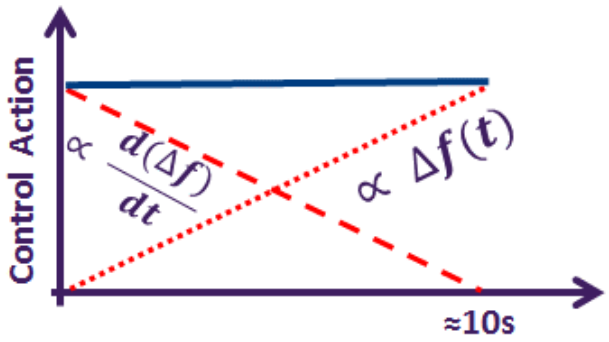


Figure 17 Schematic depiction of the control action with the hybrid strategy. The red dotted curve is the contribution due to the frequency deviation, while the red dashed curve represents the contribution of the rate of change of frequency.

The red dotted curve is the contribution to the control action due to the frequency deviation; few instants after the generation outage, the amplitude of the frequency deviation is still low and therefore the control action would be poor. The maximum response is obtained around ten seconds later, when frequency achieves the nadir; note that the deterministic controller performs exactly this control action. The control action of the hybrid controller instead benefits also from contribution of the rate of change of frequency (RoCoF); now, the RoCoF is maximum just after the generation outage and it is nil when frequency achieves the nadir. Therefore, the sum of the control actions (blue solid) due to the frequency deviation and the RoCoF allows for an approximately constant response during the whole fast transient period.

In a simple version, we consider a sampling rate of 1 Hz for the frequency evaluation. Each appliance therefore needs 1 second (2 consecutive samples) to estimate the rate of change of frequency, and one additional second to switch itself on or off. In general, a faster response is more effective. Moreover, a drastic reduction of the synchronization and a delayed payback is obtained due to the inclusion of a randomized disengagement strategy, which is triggered only once and if $(t \geq t_r)$ and $\Delta f(t) \leq 0.05\text{Hz}$. Hence, each appliance in the *on* state selects a random temperature T_{rnd} within the uniform distribution $[T_{min}, T_{max}]$; when the device reaches this temperature, its controller immediately switches its operating phase (see Figure 16b). Note that increasing t_r entails a slightly delayed onset of payback but also a faster recovery of the devices.

3.5 Case study and Results

This section deals with the set-up of the case study; simulations results are used to compare the performances of the proposed hybrid controller with those offered by three alternatives:

- *Reference case:* there is no demand side response. TCLs are treated as regular base load.
- *Deterministic controller:* the control algorithm is illustrated in section 3.4.1; the temperature thresholds update follows equation (3.10); in particular $K_f = 2$.
- *Stochastic controller:* the control algorithm is illustrated in section 3.4.2; the target temperature update follows equation (3.15); in particular $K_s = 2$. Note that the stochastic control in [31] is designed for first order device models. The thermal model used here for simulations are illustrated in (3.7).

The contribution of demand side response is tested on a simplified power system model as described by (3.19);

$$\begin{bmatrix} \Delta\dot{P}_{v1} \\ \Delta\dot{P}_{m1} \\ \Delta\dot{P}_{v2} \\ \Delta\dot{P}_{m2} \\ \Delta\dot{\omega} \\ \Delta\dot{\omega}_I \end{bmatrix} = \begin{bmatrix} -\frac{1}{T_g} & 0 & 0 & 0 & -\frac{1}{R\cdot T_g} & 0 \\ \frac{1}{T_t} & -\frac{1}{T_t} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{T_g} & 0 & -\frac{1}{R\cdot T_g} & \frac{1}{T_g} \\ 0 & 0 & \frac{1}{T_t} & -\frac{1}{T_t} & 0 & 0 \\ 0 & \frac{1}{M} & 0 & \frac{1}{M} & -\frac{D}{M} & 0 \\ 0 & 0 & 0 & 0 & K_I & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta P_{v1} \\ \Delta P_{m1} \\ \Delta P_{v2} \\ \Delta P_{m2} \\ \Delta\omega \\ \Delta\omega_I \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\frac{1}{M} \\ 0 \end{bmatrix} \cdot [\Delta P_{net}(t)] \qquad (3.19)$$

We refer the reader to [58] for a more detailed derivation of (3.19). The system model therefore consists of a linear 6th order model where we integrate the contribution of the responsive demand in $\Delta P_{net}(t) = \Delta P_L - \Delta P_{TCLs}(t)$, where ΔP_L represents the generation loss and ΔP_{TCLs} is the demand side power reduction; moreover (3.19) considers the contribution of primary control and the secondary response supplied by selected machines. We simulate a sudden loss of generation, set at 1.8 GW (maximum expected generation loss for the GB system [11]). The total

system demand is 60 GW. The parameters are chosen to qualitatively[2] match the behaviour of the GB power system as detailed by the National Grid [12]. In particular, $H = 4.5s$, $D = 1$, $T_g = 0.2s$, $R = 0.3$, $T_c = 7s$ and $K_I = 0.01$. The software used for simulations is Matlab [74].

We only consider domestic refrigeration units and we describe the temperature dynamics of these devices with the detailed fourth/second order models (3.3). Hence, we simulate 10000 individual appliances for each class (fridges, freezers and combined fridges freezers); each set is scaled up in order to approximate, respectively, 11 million fridges, 14 million freezers and 22 million fridge-freezers. The temperatures and the status of each device are randomly initialized according to the steady-state distribution; moreover, the parameters of each appliance are chosen randomly within a ±20% uniform interval around the nominal values. The aggregate power of the devices P_o is 1.37 GW. Finally, we used the values $K_f = 2$ and $K_r = 2$; t_r is set to 300 seconds.

Figure 18 shows the power consumption of the cold appliances after a generator loss event at t=1s. The hybrid control (blue), by using the RoCoF in the update of the temperature set points, permits a complete and quicker reduction of the power consumption at short notice.

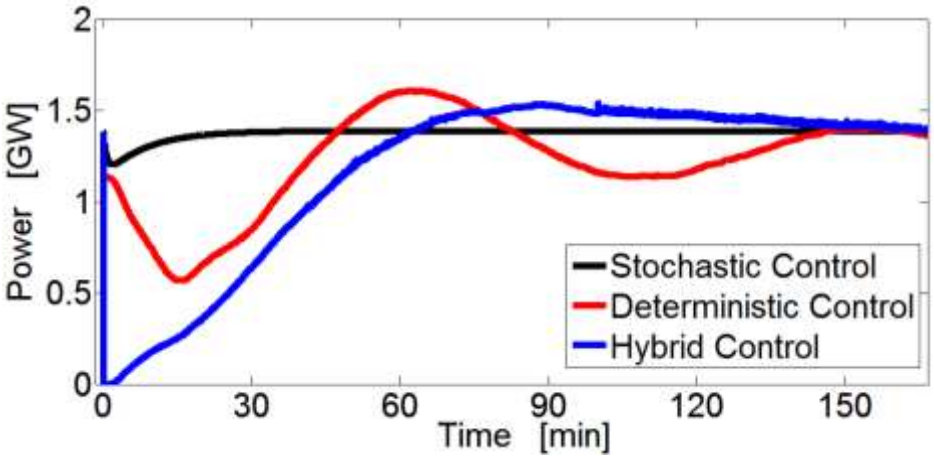


Figure 18 Time evolution of the aggregate power consumption of dynamic demand cold appliances after a 1800 MW loss with 60 GW of system demand: comparison between hybrid control (blue), stochastic control (black) and deterministic control (red).

This is not guaranteed with other strategies. In particular, the algorithm of the stochastic controller (black), clearly, does not allow for a significant power reduction. The random disengagement procedure delays the payback energy compared to the deterministic solution and

[2] Note that parameters H and D are typical values taken from [58]; T_g, R and T_t are tuned such that the corresponding frequency nadir occurs within 10s and it lies between 49.5-49.2 Hz for 1.8 GW infeed loss and for medium-high system demand. Finally, K_I is chosen to complete the generator provision of secondary response within 30 minutes as prescribed by [78].

avoids large-scale cyclic load patterns. As expected the synchronization and the payback recovery period are both suppressed with the stochastic implementation.

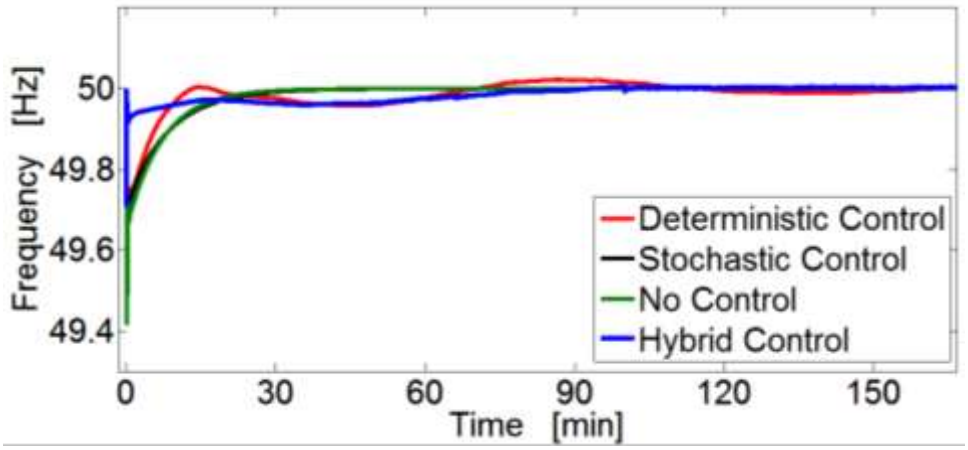


Figure 19 Frequency evolution after 1800MW loss and 60GW of system demand: comparison between the strategies (blue hybrid, black stochastic and red deterministic controller) and the reference case (green).

Moreover, Figure 19 demonstrates that controlling smarts loads with the hybrid strategy ensures the fastest recovery of frequency within a narrow range around 50 Hz. Figure 20 provides a zoom of the frequency curves on the first 15s after the generation loss; the fast and significant reduction of power consumption in case of hybrid controller is translated in a higher frequency nadir.

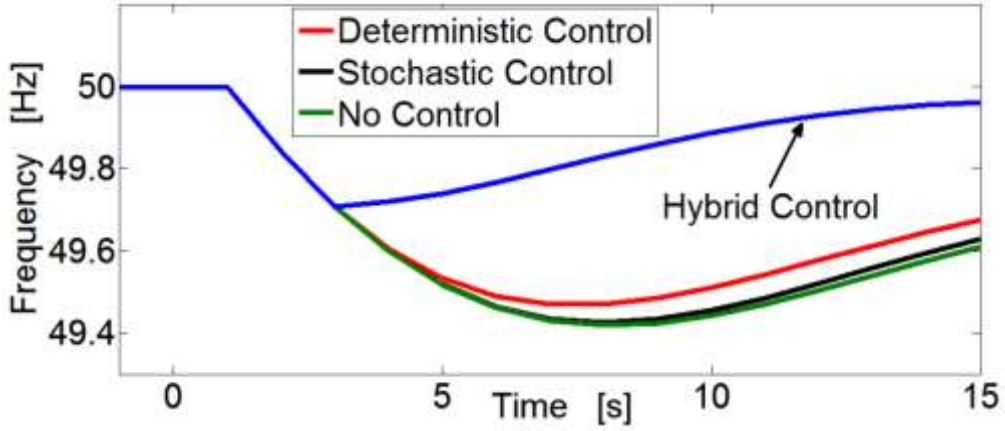


Figure 20 Zoom on frequency evolution: comparison between the hybrid control (blue), deterministic control (red), stochastic control (black) and the reference case (green).

Finally, Figure 21 also highlights the ability of dynamic demand to replace fast frequency services provided by expensive and pollutant generators. A good benchmark is obtained by comparing the time evolution of the primary and secondary control support provided by conventional generators.

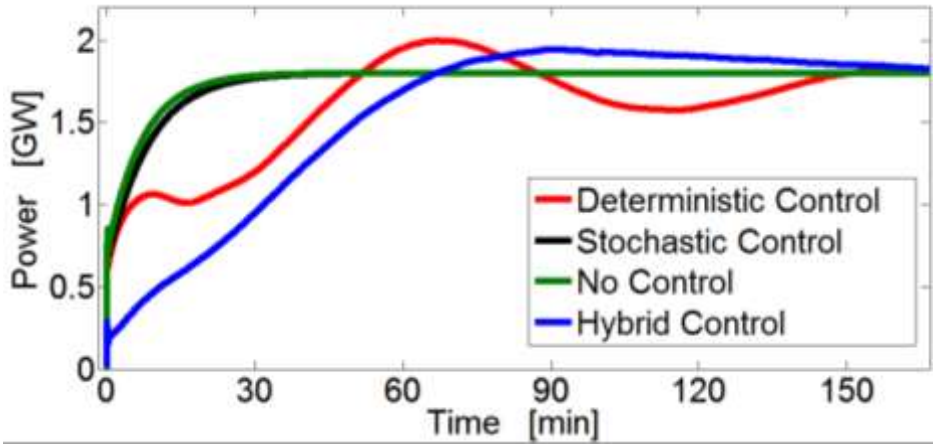


Figure 21 Power supplied by selected generators in charge of primary and secondary control. Comparison between the control strategies (blue hybrid, black stochastic and red deterministic controller) and the reference case (green) in case of 1800 MW loss and 60 GW of system demand.

There is no substantial difference between the reference case (green) and stochastic controller (black); as a consequence dynamic demand operated in this way is not economically nor technically convenient. Considering the hybrid controller (blue), concrete benefits in the provision of primary response support are highlighted as the curve starts from a much lower value. Moreover, up to slightly more than one hour after the generator outage, the hybrid curve is always below the reference case, so less conventional generation is required. The initial ramp rate conventional of generators decreases from ≈115 MW/min to ≈23MW/min. This is the key outcome of the hybrid strategy, as the amount of fast but perhaps expensive and/or pollutant generation for frequency support can be reduced, leading to economic and environmental savings. Furthermore, the payback effect does not undermine these benefits since the extra generation required is limited and occurs more than an hour after the initial event, a sufficient time to reschedule the system in a more efficient way.

3.6 Conclusions

This chapter introduces the thermal dynamic models that describe the temperature dynamics of various refrigeration loads in the domestic and in the commercial sector. Detailed second order models have been used for domestic refrigerators, freezers and combined fridge-freezers (fourth order model). Effective first order models have been introduced for commercial units because of complexities in finding details for high-order dynamic models. Furthermore, in this chapter we investigated initial controller designs [28, 29] for frequency response with TCLs.

Although such controllers provide an effective initial response to frequency deviations, careful analysis of their long-term response shows a tendency for devices to synchronize their cooling cycles. This interferes with the diversity of demand on the network, and may result in self-reinforcing frequency oscillations.

The stochastic controller designed in [31] improves the robustness of the control action. The aim of this strategy is to eliminate the payback phase, but in doing so it prolongs the time it takes for appliances to regain their steady state temperature, and limits the ability to implement rapid load changes; this is an important indicator of the demand side support to system frequency control. At the end of the chapter, we proposed a hybrid control strategy for thermostatic appliances for the provision of frequency services. By monitoring the rate of change of frequency, the controller enables a rapid response after the loss of a generating unit. Furthermore, the addition of a randomized disengagement strategy both postpones the payback and drastically reduce the synchronization effect observed in simpler control schemes. We have investigated the performance of the hybrid controller by using a second order model for domestic refrigerators and domestic freezers and a fourth order model for domestic combined fridge-freezers, coupled to a simple power system model. Comparisons with the deterministic and stochastic controllers are included, confirming that the hybrid controller addresses previously identified shortcomings.

3.6.1 Extensions and links with the next chapters

This chapter pointed out some of the difficulties linked to the accurate control of thermostatically controlled loads mainly due to the peculiar *on/off* operating mode that characterize these devices. Despite all the benefits provided by the hybrid controller, this design represents an ad-hoc solution for the synchronization issue and moreover it does require careful tuning of parameters for specific scenarios. This strategy still does not provide full controllability of the power profile over a large time scale (specifically after the initial response). The achievement of full and robust controllability of a large population of thermostatic loads is the main topic of the next chapter.

Chapter 4: Advanced Control for Thermostatically Controlled Loads and its Impact on System Frequency Control

Abstract

This chapter deals with a novel advanced framework for the control of thermostatically controlled loads; this control architecture permits precise modulation of the aggregate power consumption of a large heterogeneous population of devices through stochastic control. The main outcomes of this methodology involve the thermostatic loads' ability to track arbitrary power profiles (within limits); this result creates the room for an interesting variety of demand response applications without the need for hard constraints on the communication system. In addition, the control method is designed for decentralised implementation and it guarantees the accurate respect of temperature limits for each individual appliance. Sections 4.2-4.7 are a summary of the main outcomes described in [56]; the paper is joint work with Simon Tindemans. The contribution to that work from the author of this thesis regards (a) the definitions of the system-level and device-level requirements, (b) the descriptions of devices models, (c) the definition of the control challenges and (d) the active discussion on high-level framework of the control architecture. Moreover, in Section 4.8 we present new results and applications. The high-level properties enable the design of two implementations for enhancing the system frequency control; the first, considers the tracking of pre-programmed power curve in response to frequency events. The second instead makes the TCLs aggregate response proportional to the system frequency evolution. This chapter also evaluates the impact of both these two implementations on the system frequency control under a future low-carbon scenario.

4.1 Introduction

In the previous chapters, we have introduced a particular set of appliances, the thermostatically controlled loads (TCLs) and we described the design of the thermal dynamic models that describe the temperatures' evolution of these appliances. Moreover, the regular operation of an individual thermostatic load has been illustrated to highlight the peculiar *on-off* operating characteristic of these devices. The pioneering idea proposed by Schweppe in 1979 [39] demonstrated the ability for these appliances to positively participate to the system frequency control. The overall concept is to exploit the TCLs insensitivity to relatively short-term temperature fluctuations, creating a flexible distributed load shedding in order to quickly reduce the generation-demand imbalance and therefore arrest the consequent frequency deviation.

The fundamentals of the work by Schweppe have been reconsidered in later works [28, 29, 40]; these deterministic implementations discussed in Section 3.4.1 are heuristic modifications of the regular temperature-threshold thermostatic controller (see Section 3.2). Although these implementations offer positive ability to limit the initial frequency deviation, a more careful study of TCLs' long-term response reveals the tendency for the devices to synchronize their operating cycles [29]. The design proposed by Angeli and Kountouriotis [31] made use of random switching in order to cancel the synchronization drawback; this work represents the first attempt to guarantee the controllability the long-term TCLs response. The controller is able to modify the properties of the steady state distribution but this employs a slow control over the temperature distribution of a cluster of TCLs. A tangible result is the elimination of the payback period (i.e. no extra power absorption), but in doing so the time it takes required by the devices to reach their steady state condition (in temperature) drastically increases. However, the stochastic controller developed in [31] only partially solved the issues arisen from the deterministic implementations.

In order to overcome the issues linked to these frameworks we proposed a hybrid version that drastically reduces the synchronization by means of a simple disengagement strategy. In doing so, the payback period becomes smoother and sufficiently delayed to not represent an issue for the generators that provide reserve service; however, this method still does not provide the full control over the power profile beyond the initial response. The complexities encountered in the assessment of initial controller designs for TCLs provision of frequency support suggest that it is not

straightforward to design a control algorithm for demand side application that simultaneously satisfies short-term and long-term precise requirements. [49]; the difficulties are mainly due to the typical TCL operating cycle characterized by only two power states (*on* and *off*). A multitude of devices should therefore be controlled in harmony to provide a secure service to the network. Hence, in this chapter we start our analysis with the identification of four main features for demand side response applications with TCLs that a controller has to provide.

1) Accurate control across a range of time scales. Thermostatic loads have to react to frequency events within a narrow time window (seconds) and conduct a precise recovery phase to the regular operation, defined by a steady state power consumption and a steady state average temperature. Although the complete execution of a demand side response service for frequency control would not exceed a one hour commitment, wider time windows would be required for other demand side applications such as energy arbitrage.

2) Freedom to design complex responses. The TSOs or DSR aggregators should have the ability to tailor responses in accordance with both technical and commercial requirements. Clearly, the design should take into account the physical capabilities of the TCLs; individual classes of TCLs could deliver significantly different responses due to physical differences between them[3].

3) Reliable distributed response. The implementation of a real-time communication between a central controller and individual devices (millions of TCLs) implies the need for a high-cost communications infrastructure; moreover even facing these cost does not represent a warranty of success as these infrastructure will be sensitive to disturbances. It is therefore desirable to resort to decentralised frameworks based on locally available control signals (frequency, time).

4) Satisfy per-appliance quality of service constraints. The TCLs' participation to demand side programs should not affect the primary function of the TCLs: maintaining the controlled temperature around a set point value. Therefore, a well-designed controller must deliver the response without violating the quality of service requirements (e.g. temperature limits) for each individual appliance.

[3] In Chapter 5 we will focus on how physical characteristics of individual classes of devices affect the final ability to provide a particular ancillary service.

The available methods discussed in the previous chapter and other approaches developed during recent years address one or more of these challenges, but not all simultaneously. In fact, the traditional temperature-threshold controller can be modified to permit the tracking of particular power profiles by shifting the temperature deadband in real time [41, 42, 33]. The underlying assumption that characterize this approach is the availability of a real-time communication structure [34]. However the limitation of this approach lies in the use of a single control parameter (individual temperature set point); in fact, the variety of response actions would be restricted by hard device-level temperature limits. Furthermore, the proposed implementations build on a several analytical and numerical approximations [75] that complicate the analysis and design of response actions.

In this chapter, we focus on the main outcomes and properties of a novel control strategy (Sections 4.2-4.7) for TCLs that, for the first time, addresses all four challenges listed above. Furthermore, in Section 4.8 we include additional results regarding the impact on the network frequency response of controlling TCLs with this novel control algorithm. The core of the control architecture developed deals with the ability to track arbitrary power profiles (within limits) using independent actions of a heterogeneous cluster of TCLs. This result triggers a variety of demand response application without the need for a high-cost and unreliable communication infrastructure. Although the control architecture is suitable for complex designs, we introduce here a simple controller for which the associated control actions is derived analytically. This choice enables the explicit quantification of the contribution that aggregate thermostatic loads can provide in terms of power and energy levels. This analysis, carried out considering devices characterized by identical parameters (homogeneous population) is extended to formulate envelope constraints for the aggregate flexible contribution of heterogeneous devices.

Provided this ability of the advanced controller to accurately command in a decentralized fashion a large population of thermostatic loads, we compare two different implementations that tailor different aggregate power response; both methods fit controller's framework. The first control implementation strictly follows the settings in [56] and considers the tracking of pre-programmed power curve. The second method instead makes the TCLs response a linear function of the system frequency and its rate of change. Initial temperature deadband controllers ([29, 28] and others) could not guarantee such a linear relation during the whole transient evolution of the system frequency; in fact, these controllers induce an initial power reduction that is only

approximately proportional to the initial frequency drop. When the system frequency recovers, the long term response shows a tendency for devices to synchronize their cooling cycles. Four different versions for the frequency linear controller are proposed and tested. We compare the performances of these controllers on a network model that mimics the dynamics of the future low carbon GB power system. Computer simulations test the increased ability of the network to integrate more renewable generation when TCLs are controlled with the methods proposed.

The chapter is organised as follows. In section 4.2 we present the general framework of the control architecture; the basic idea is that TCLs are controlled in a fully decentralised fashion and each individual unit can independently target a reference power profile. Hence we discuss the further benefits that 'semi-autonomous' control offer compared to 'fully autonomous' control. Section 4.3 explains in depth the two-level structure of the control framework. Section 4.4 deals with the first level, the ensemble problem, where the precise tracking of the reference power profile is guaranteed by analysing the dynamic evolution of the devices' temperature. The results is that the net heating rate $v(T,t)$ can be smartly used as control parameter. Hence, the aggregate power consumption of TCLs can be driven the population-averaged rate of heating or cooling. Moreover, Section 4.5 applies the control properties to the linear first order thermal models that describe the individual appliances. For such a class of models we analytically propose the relation between temperature and power consumption. This result is used to infer high level limits on the flexibility of appliances, expressed in terms of temperature (energy) and (instantaneous) power consumption. In section 4.6 we present an straightforward implementation of the control structure; the heating rate $v(T,t)$ used as control function is assumed to be a linear function of T; this feature permits the analytical derivation of the controller's properties. Then, Section 4.7 briefly touches the second step of the control problem: the device-level problem and recalls the results in [56]. In Section 4.8 we assess the impact of two different implementation of the control framework on the system frequency control and, finally, Section 4.9 concludes the chapter a provides the links to the further chapters.

4.2 General framework

The methodology developed in [56] proposes to control the aggregate power consumption of TCLs in a decentralised fashion; each appliance a can independently track a relative power profile $\Pi(t)$. The individual power consumption $P^a(t)$ is modulated in such a way that:

$$E[P^a(t)] = P_o^a \cdot \Pi(t) \tag{4.1}$$

where P_o^a is its steady state power consumption. The notion of expectation used in this expression is with respect to the statistical ensemble of all possible initial conditions of appliance a (temperatures, states) and – for stochastic controllers – all possible control sequences. The reference profile $\Pi(t)$ is identical for all the devices; moreover, by construction, a response action starts and ends with $\Pi(t)$ = 1 (steady state). Provided that the response of each device respects (4.1), it is easy to be convinced that $\Pi(t)$ also controls the aggregate power consumption $P^{total}(t)$:

$$E[P^{total}(t)] = \sum_a E[P^a(t)] = \sum_a P_o^a \cdot \Pi(t) = P_o^{total} \cdot \Pi(t) \tag{4.2}$$

where P_o^{total} is the steady state aggregate power consumption. It is worth pointing out that because the devices are statistically independent of each other, the relative deviation from the expectation will decrease approximately as $1/\sqrt{N}$, where N is the number of appliances. The aggregate power of TCLs may be described by a binomial distribution with parameters $n = N$ and $p = \pi$, the number of devices and the probability for each device of being on (i.e. the duty cycle), respectively. For large N the binomial distribution can be approximated by a normal distribution with mean $N \cdot \pi$ and standard deviation $\sqrt{N\pi(1-\pi)}$; it follows that the standard deviation of the generalized distribution (dividing for the number of appliances) will be proportional to $1/\sqrt{N}$. For large N we may therefore assume

$$P^{total}(t) \cong P_o^{total} \cdot \Pi(t) \tag{4.3}$$

We now proceed to summarise the high-level control structure; the architecture starts with the demand 'response designer' that determines the specific demand response profile. This role can be played by the TSO, but depending on the market regulations this task can also be fulfilled by demand response aggregators.

The desired relative power profile $\Pi(t)$ could dynamically depend on the network condition. A high-reliable communication infrastructure would be required to transmit to the devices such a curve in real time; however this solution would drastically increase the investment cost for these applications and still it would not guarantee the insensitivity to disruptions. Instead, following the approach in [56] the system operator can design a *power response model* that is broadcast to the devices. The power response model establishes how the relative power profile $\Pi(t)$ can be constructed as a function of a locally available signal, such as the grid frequency or the current time. This way, each device has the ability to build the desired $\Pi(t)$ independently. For instance, a linear relation between the power consumption of the appliances and the network frequency can be imposed.

In a basic application the power response model may be already incorporated in the appliance during production phase. This approach reflects a fully autonomous operation and it is suitable for to the provision of primary response control. However, this operating mode does not consider the potential significant benefits originated by a power response model periodically updated. Modifying the response model on a long-term basis (hours or days) allows the system operator to optimally match the scheduled response services and the actual power system condition (level of wind/demand load/ generators available). Moreover, it would enable the provision of energy arbitrage services that reduce peak load, generation cost and/or ramp constraints. This operating mode can be called 'semi-autonomous', as the modifications of the regular response program are made on a time scale that is sufficiently larger than the one related to the response services (seconds - minutes). To achieve this, there is no need for real-time communication (as for centralised control) so the latency of the communication link is not critical; it is sufficient for the channel to be sufficiently reliable. The roll out of the smart metres accommodates these characteristics. It is therefore well suited to support the proposed semi-autonomous control of TCLs.

The scheduling of demand response services requires that the desired curve $\Pi(t)$ can be actually followed by the appliances without any risk of infringing device-level temperature constraints. Two consequences arise from this premise. The former forces the system operator to be aware of the aggregate capability of the participating devices, especially their ability to modulate their power levels and temperatures. The latter imposes the need for effective means to evaluate the compatibility of device parameters and any desired power profiles. For the controller in analysis, a set of aggregate parameters and conditions is sufficient to ascertain the compatibility with a

proposed power profile $\Pi(t)$. (See Section 4.7.1). The following section focuses on the design of the local controller for a *single* device that controls its power consumption with respect to to the target relative power profile $\Pi(t)$.

4.3 A two- level controller

This section offers a more detailed explanation of the advanced stochastic controller, in this context particular attention is devoted to the description of the intrinsic structure of the control philosophy. Device-centric approaches rapidly lead to undesirable synchronisation effects [29, 28]. System centric approaches, on the other hand, often compromise device performance causing large temperature fluctuations [41, 42], along with excessive switching rates and slow initial response and recovery [31]. A two-stage controller for thermostatic loads is presented in [56]; it enables an efficient provision of frequency support services while reducing the impact on the quality of the service itself. The first step (ensemble problem) works following a temperature probability distribution and guarantees that a specified average power curve is tracked without exceeding temperature thresholds of individual devices. The second stage (device control problem) is a stochastic device-level controller that produces the necessary modifications in the probability distribution. Specifically this phase computes the switching rates r_{off}^{on} (*on* → *off*) and r_{on}^{off} (*off* → *on*) that are additional to the regular deterministic switching, $T_{low}(t)$ and $T_{high}(t)$.

Considering that the thermal appliances are characterized by an internal state that is completely defined by the temperature, they can therefore be modelled by a first order ODE. Cold thermostatic loads like refrigerators are used as reference loads although the same ODE model can simply be applied to other TCLs such as air conditioning units and resistive space heaters. Every appliance is assumed to have an *on* or *off* state, but extensions to devices with fractional power states are also conceivable. The temperature evolution of a single device $T^a(t)$ in state s is defined by the differential equation

$$\frac{dT^a(t)}{dt} = \begin{cases} v_{on}(T^a(t)), & when\ s = on \\ v_{off}(T^a(t)), & when\ s = off \end{cases} \tag{4.4}$$

where $v_{on}(\cdot) < 0$ is the (active) cooling rate and $v_{off}(\cdot) > 0$ the (passive) heating rate. Section 3.2 illustrates the typical operation of a refrigerator device using the temperature deadband

$[T_{min}, T_{max}]$ around the set point T_{set}. The traditional approach suggests to enable the demand response from TCLs in order to extend this control strategy by shifting the upper and lower temperature bounds in unison ([29, 28, 41] and others). In the context proposed in [56], both deterministic and stochastic switching are used, hence:

- *Variable thresholds.* The appliance always switches off when the lower temperature threshold $T_{low}(t)$ is reached and switches on when the upper temperature threshold $T_{high}(t)$ is reached. This guarantees that the temperature of an individual appliance never exceeds the interval $\left[T_{low}(t)\ , T_{high}(t)\ \right]$.

- *Stochastic switching.* For the intermediate temperature $T \in \left[T_{low}(t)\ , T_{high}(t)\ \right]$ the state jumps are forced by stochastic switching rates r_{off}^{on} (*on* → *off*) and r_{on}^{off} (*off* → *on*). If these rates equal zero, the control architecture will embody the conventional set point controller.

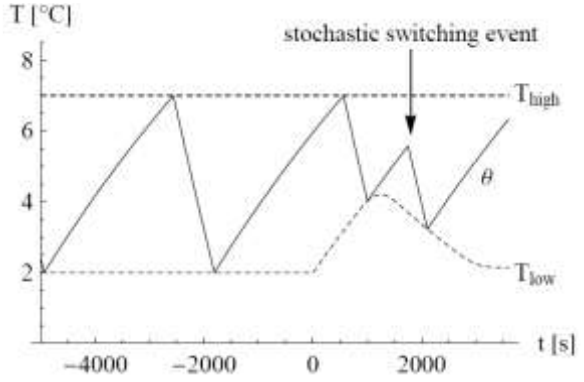


Figure 22 Illustrative temperature trace of a single refrigerator controlled by a hybrid threshold-stochastic controller.

Figure 22, where the temperature trace of a single refrigerator is plotted, depicts the control strategy aforementioned. For $t < 0$ the controller performs like a regular deadband controller and the temperature fluctuates between the constant thresholds $T_{low}(t) = T_{min}$ and $T_{high}(t) = T_{max}$. After $t = 0$ the lower temperature limit $T_{low}(t)$ varies progressively and occasional stochastic switching augment the threshold switching.

4.4 The ensemble problem

The aim of the ensemble problem is to construct an appropriate population-level response for tracking a defined relative power profile $\Pi(t)$. This first order ODE (4.4) elicits a probability density function $f(T,t)$ of devices; this distribution results to be function of both temperature and time. Moreover it results straightforward to explain this distribution as the probability density of temperatures for an appliance with unknown initial conditions.

$$\frac{\partial}{\partial t} f(T,t) = -\frac{\partial}{\partial T}[v(T,t) f(T,t)] \tag{4.5}$$

with $v(T,t)$ the average heating rate of all the appliances at temperature T, which instructs the 'flow' of appliances along the temperature axis. Equation (4.5) points out that the evolution of the temperature distribution $f(T,t)$ is fully defined by the heating rate $v(T,t)$. Furthermore, considering the probability density of a population of TCLs over temperature T that, at time *t*, are in an *on* $\left(f_{on}(T,t)\right)$ or *off* $\left(f_{off}(T,t)\right)$ phase, it is obtained

$$f(T,t) \equiv f_{on}(T,t) + f_{off}(T,t) \tag{4.6}$$

$$v(T,t) \equiv \frac{v_{on}(T,t) f_{on}(T,t) + v_{off}(T,t) f_{off}(T,t)}{f_{on}(T,t) + f_{off}(T,t)} \tag{4.7}$$

Obviously, $v(T,t)$ has to respect that $v_{on}(T,t) \leq v(T,t) \leq v_{off}(T,t)$. These considerations introduce the main contribution of this control architecture: the heating rate $v(T,t)$ is used itself as a control parameter. Considering an individual TCL, at each instant its compressor is either *on* or *off* thus the consequent power absorption is either P_{on} (the maximum power consumption) or nil. However, using (4.6) and (4.7), we are able to calculate the fractional power consumption $\pi(t) = E[P(t)]/P_{on}$ as

$$\pi(t) = \int_{-\infty}^{\infty} f_{on}(T,t)\,dT = \int_{-\infty}^{\infty} \left(\frac{v_{off}(T) - v(T,t)}{v_{off}(T) - v_{on}(T)}\right) f(T,t)\,dT \tag{4.8}$$

where $\pi(t) \in [0,1]$ and the relative power consumption are related by

$$\Pi(t) = \frac{\pi(t)}{\pi_o} \tag{4.9}$$

and π_o is the steady state fractional power consumption, equal to the duty cycle.

At this first stage (ensemble problem) the steady state distribution $f(T,0)$ is calculated and defines the starting point. Afterwards, each device computes a suitable population-level response for tracking the target relative power curve $\Pi(t)$; this action is achieved by choosing the heating profile $v(T,t)$ in accordance with the target power curve $\Pi(t)$ and equation (4.8). The information required are only based on the devices' thermal models and their temperature limits. This aspect will be analysed in the following sections.

4.5 Linear thermal model

This section shows that the use of linear first order models for individual appliances results in interesting population-level properties. Moreover, it is worth pointing out that the results described in the remainder of this section do not depend on the specific details of the controller, and thus can be applied to any *on-off* controller for first linear thermal models. We recall the first order thermal model (3.7) introduced in Section 3.3.2; the heating/cooling rates ($v_{on}(T)$ and $v_{off}(T)$, respectively) turn into:

$$v_{on}(T) = -\frac{1}{\tau}(T - T_{on}) \tag{4.10a}$$

$$v_{off}(T) = -\frac{1}{\tau}\left(T - T_{off}\right) \tag{4.10b}$$

4.5.1 Temperature and power consumption

The fractional power consumption $\pi(t)$ as function of $f(T,t)$ and $v(T,t)$ is determined by inserting (4.10) into (4.8)

$$\pi(t) = \int_{-\infty}^{\infty} \left(\frac{T_{off} - T - v(T,t)\tau}{T_{off} - T_{on}} \right) f(T,t)dT = \frac{1}{T_{off} - T_{on}} \left[T_{off} - \bar{T}(t) - \tau \cdot \bar{v}(t) \right] \tag{4.11}$$

where $\bar{T}(t)$ represents the average temperature and $\bar{v}(t)$ is the average rate of heating at time t. At the steady state it results that $\bar{v} = 0$ and this gives rise to following relation between the steady state average temperature and $\bar{T}_o$ and the duty cycle π_o:

$$\pi_o = \frac{T_{off} - \bar{T}_o}{T_{off} - T_{on}} \tag{4.12}$$

By definition $\bar{v}(t) = d\bar{T}(t)/dt$, therefore equation (4.11) can be interpreted as a differential equation for $\bar{T}(t)$. Considering the relative power consumption $\Pi(t) = \pi(t)/\pi_o$ it becomes

$$\frac{d\bar{T}(t)}{dt} = -\frac{1}{\tau}\left[\bar{T}(t) - T_{off} + \left(T_{off} - \bar{T}_o\right)\Pi(t)\right] \tag{4.13}$$

The population average temperature $\bar{T}(t)$ is proved to evolve as that of a single TCL characterized by a variable relative cooling rate $P(t)$=$\pi_o P_o \Pi(t)$. The solution to the differential equation (4.13) is given by:

$$\bar{T}(t) = T_{off} + \left(\bar{T}(t_0) - T_{off}\right)e^{-\frac{t-t_0}{\tau}} - \frac{1}{\tau}\left(T_{off} - \bar{T}_o\right)\int_{t_0}^{t} \Pi(t')\, e^{-\frac{t-t'}{\tau}} dt' \tag{4.14}$$

taking the limit $t_0 \to -\infty$ we rearrange (4.14) as

$$\bar{T}(t) = T_{off} - \frac{1}{\tau}\left(T_{off} - \bar{T}_o\right)\int_{-\infty}^{t} \Pi(t')\, e^{-\frac{t-t'}{\tau}} dt' \tag{4.15}$$

This equation give rise to an interesting property: the average temperature is determined by an exponentially smoothed reference power profile $\Pi(t)$.

4.5.2 Energy and power constraints

The quality of the service of thermostatic load depends on the ability to keep a compartment at a set-point temperature. The provision of any ancillary service cannot excessively affect this task; hence, defined the operating interval $[T_{min}, T_{max}]$, the temperature of each individual appliance has to lie within this range at all times. This constraints represents a discriminating factor for the desired power profiles $\Pi(t)$, both in terms of energy and power.

1) Capacity for energy services: The energy constraint involves the ability to maintain a reduced power consumption, so that the power system 'borrows' the thermal energy from the devices. This valuable energy tank is fully depleted when each device reaches the upper temperature threshold T_{max}, so that $\bar{T} = T_{max}$. On the other side, the storage capacity is fully available when each TCL is at its lower temperature limit T_{min}, so that $\bar{T} = T_{min}$. An optimal controller is able to exploit the full temperature range; therefore the average temperature lies within maximum and minimum temperature boundaries

$$T_{min} \leq \bar{T}(t) \leq T_{max} \tag{4.16}$$

Such a temperature limitations is converted in a constraint on the reference power profiles. Using equation (4.15) we obtain,

$$1-\left(\frac{T_{max}-\bar{T}_o}{T_{off}-\bar{T}_o}\right) \leq \frac{1}{\tau}\int_{-\infty}^{t} \Pi(t')\, e^{-\frac{t-t'}{\tau}} dt' \leq 1+\left(\frac{\bar{T}_o-T_{min}}{T_{off}-\bar{T}_o}\right) \tag{4.17}$$

Assuming a constant value for $\Pi(t)=\Pi_{sustained}$ we obtain the following inequality for power levels that can be maintained indefinitely without the risk of violating (4.16):

$$1-\left(\frac{T_{max}-\bar{T}_o}{T_{off}-\bar{T}_o}\right) \leq \Pi_{sustained} \leq 1+\left(\frac{\bar{T}_o-T_{min}}{T_{off}-\bar{T}_o}\right) \tag{4.18}$$

Note that equation (4.15) suggests that operating at a constant power level eventually results in a constant average temperature. In particular, we point out a general result: it is always possible to asymptotically recover the steady state temperature by fixing the operating level at $\Pi = 1$. Therefore, it is not a strict requirement to absorb extra power consumption in order to recover the steady state energy condition after an initial demand reduction. This convenient property was used successfully in [31] to avoid fluctuating power levels during the recovery phase. Nevertheless, a payback phase characterized by a 'payback' period may be useful to temporarily boost power consumption in order to speed up the average temperature recovery.

2) Instantaneous power consumption: It is not possible to derive similar generic bounds for the instantaneous power consumption without referring to a controller implementation. In principle, the use of the full range of fractional power levels between $\pi(t)=0$ and $\pi(t)=1$ is guaranteed simply by randomly selecting a fraction $\pi(t)$ of appliances that will be in the *on* state. Clearly some devices cannot be switched on and off as it would immediately result in a violation of the quality of service constraints. However, the fraction of appliances for which this applies is usually negligible. Therefore, without further specifications on the controller implementations it is possible to state the following approximate constraints:

$$0 \lesssim \Pi(t) \lesssim \frac{1}{\pi_o} \tag{4.19}$$

In section 4.6.2 we will determine instantaneous power bounds for the specific controller developed in the next section.

4.6 Linear controller

In section 4.4 we computed the generic bounds for the aggregate demand response abilities of appliances described by means of linear thermal models. Hereafter these results will be applied to define a particular controller implementation. The linear controller in [56] does not fully exploit the flexibility of the devices, but its simple functional form results in an analytically tractable control problem. This property makes it an attractive option as an illustrative example. Furthermore, the linear controller has two advantageous aspects:

1) It automatically degrades to the traditional set point controller at steady state

2) It does not allow any TCL to overcome their maximum temperature T_{max}

4.6.1 The control parameter

In order to modulate the power consumption, the heating rate profile $v(T,t)$ (the control function) is parameterised to obtain a control *parameter*. Equation (4.121) suggests the choice of $v(T,t)$ as a linear function of T; thus we infer:

$$v(T,t) = \frac{\beta(t)}{\tau}(T - T_{max}) \tag{4.20}$$

The heating/cooling rate of the appliances' population is thus determined by means of the dimensionless control parameter $\beta(t)$, with the convention that $\beta(t) < 0$ is associated to a net heating effect (reduced power consumption) and $\beta(t) > 0$ to a net cooling effect (increased power consumption). Because $v(T_{max},t) = 0$ for any choice of β, it is observed that $T_{high}(t) = T_{max}$ at all times (as $\frac{dT_{high}(t)}{dt} = v\big(T_{high}(t),t\big)$). As a consequence, the TCLs' temperatures will not go over T_{max}, reflecting the quality of service constraints. Unlike the upper temperature limit, the lower temperature limit $T_{low}(t)$ is dynamic under this implementation of the controller as it is conceived for demand reduction. Its evolution can be always calculated through the solution of the differential equation $\frac{dT_{low}(t)}{dt} = v(T_{low}(t),t)$ with the initial condition $T_{low}(0) = T_{min}$; hence,

$$T_{low}(t) = T_{max} - (T_{max} - T_{min})e^{\int_0^t \frac{\beta(t')}{\tau}dt'} \tag{4.21}$$

substituting (4.20) in (4.11) and solving for $\beta(t)$ we obtain the control parameter as a function of the ensemble-averaged relative power consumption $\Pi(t) = \pi(t)/\pi_0$:

$$\beta(t) = \frac{\Pi(t)\left(T_{off} - \bar{T}_o\right) - \left(T_{off} - \bar{T}(t)\right)}{T_{max} - \bar{T}(t)} \tag{4.22}$$

Together equations (4.15) and (4.22) offer a way to define the mean temperature $\bar{T}(t)$ that results from a specified power profile $\Pi(t)$, and the value of $\beta(t)$ that is required to target this power profile. It is worth noting that the response depends only on the constant system parameters τ, T_{max}, T_{on} and T_{off}. As a result, the controller is capable of solving the ensemble control problem without resorting to numerical integration of PDEs, which is particularly attractive for embedded applications. Figure 23 depicts the evolution of the average temperature $\bar{T}(t)$ (B) and the control parameter $\beta(t)$ (C) for a specific relative power curve $\Pi(t)$ (A) that reduces the devices' power consumption to 50% of the steady state level and maintains this for 15 minutes.

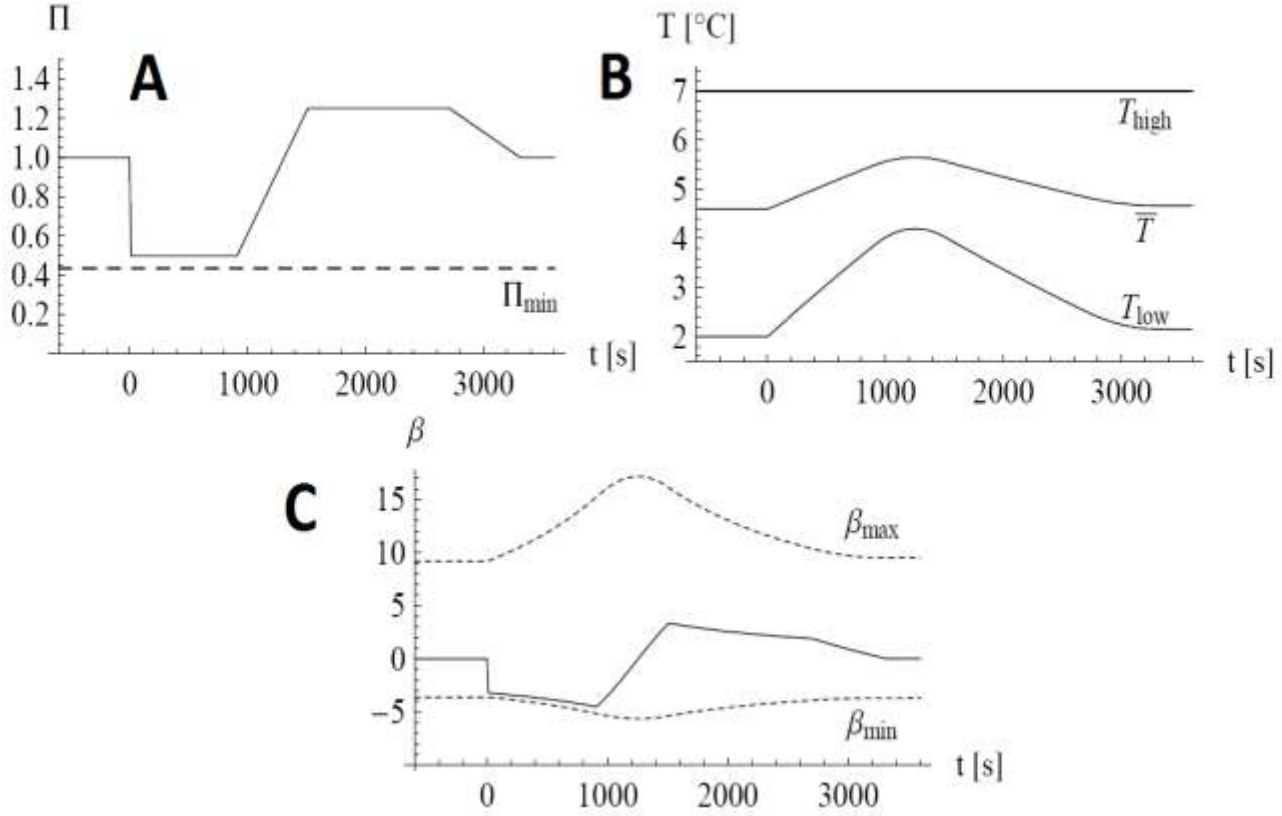


Figure 23 (A) reference relative power curve and the lower limit given by (4.29). (B) Minimum (see (4.21)), mean (see (4.15)) and maximum temperatures of the ensemble. (C) Value of the control parameter plotted alongside upper and lower limits representing the constraints in (4.22).

A tiny increase in power consumption (payback) is used to recover the steady state (i.e. the initial condition) almost exactly by 55 minutes after the initial power drop.

4.6.2 Energy and power constraints

The linear controller described by equation (4.20) has been chosen for its analytical tractability; the analysis derived focused on demand side reduction applications with a consecutive recovery period. This approach prevents the control framework from fully exploiting the physical capacity of the devices. Hence, we now study the energy and power constraints imposed by the linear controller, and compare these with the more general bounds for TCLs introduced in Section 4.5.2. It is worth pointing out that extensions to more sophisticated control implementations that also allow for a net energy absorptions from the power system. In particular in Chapter 5 we will combine the reverse scheme for demand side increase to the framework for demand reduction that we are now presenting; this action has the effect of increasing the energy bounds.

1) Energy: Let us consider the steady state condition as the starting point of the analysis; the appliances' temperatures are distributed between T_{min} and T_{max}; therefore, any initial increase in power consumption $\Pi(t) > 1$ would push the temperature of some of them below T_{min}. This issue should be actually avoided considering simple refrigerators as accidental freezing would be dangerous. In order to guarantee that no appliance exceeds the temperature range $[T_{min}, T_{max}]$ the average temperature excursion have to be limited between

$$\bar{T}_o \leq \bar{T}(t) \leq T_{max} \tag{4.23}$$

As consequence of this, the energy bounds in (4.18) are restricted to

$$1 - \left(\frac{T_{max} - \bar{T}_o}{T_{off} - \bar{T}_o}\right) \leq \frac{1}{\tau}\int_{-\infty}^{t} \Pi(t')\, e^{-\frac{t-t'}{\tau}} dt' \leq 1 \tag{4.24}$$

The comparison of equation (4.24) with (4.18) shows that the controller is optimal with respect to available energy for demand reduction. However, although the controller does not allow for initial power increase $\Pi(t) > 1$, such a situation is no longer obstructed during the payback period that can comes after an initial demand reduction (i.e. when $\bar{T}(t) > \bar{T}_o$).

2) Power: The range of possible instantaneous power levels depends on the acceptable range of $\beta(t)$. The more the devices' temperature deviate from T_{max}, the stronger the control action will be. The range of β is limited by those values at which all devices at $T_{low}(t)$ (the coldest appliances at time t) are forced to be either *on* or *off*. Figure 24 graphically illustrates the available range. In particular, the minimum and maximum values of $\beta(t)$ are obtained by solving

$$v_{off}(T_{low}(t)) = \frac{\beta_{min}(t)}{\tau}(T_{low}(t) - T_{max}) \tag{4.25a}$$

$$v_{on}(T_{low}(t)) = \frac{\beta_{max}(t)}{\tau}(T_{low}(t) - T_{max}) \tag{4.25b}$$

This results in the following constraint for $\beta(t)$:

$$-\frac{T_{off} - T_{low}(t)}{T_{max} - T_{low}(t)} \leq \beta(t) \leq \frac{T_{low}(t) - T_{on}}{T_{max} - T_{low}(t)} \tag{4.26}$$

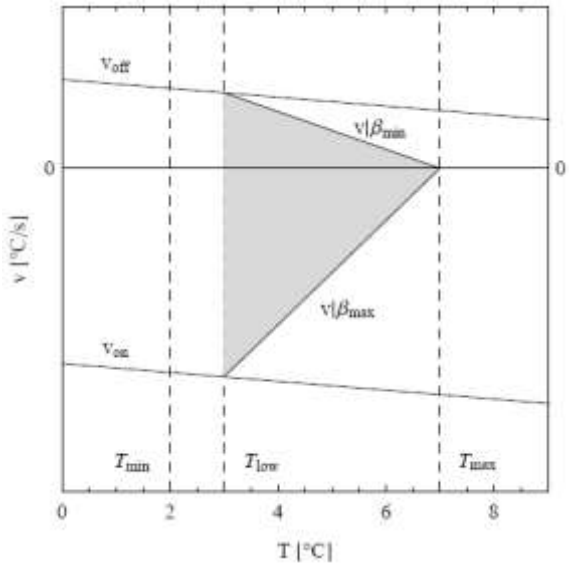


Figure 24 Graphical construction of the physical constraint on the control parameter $\beta(t)$. Appliance temperature are distributed between $T_{low}(t) \geq T_{min}$ and T_{max} and their heating/cooling rates are bounded by v_{off} and v_{on}. The permissible range of $v(T,t) = \frac{\beta(t)}{\tau}(T - T_{max})$ is shown in grey.

The constraint on the control parameter $\beta(t)$ can be translated into a limitation on the instantaneous power consumption by using (4.22) to substitute $\beta(t)$ and using (4.21) and (4.27)

$$\bar{T}(t) = T_{max} - (T_{max} - \bar{T}_o)e^{\int_0^t \frac{\beta(t')}{\tau}dt'} \tag{4.27}$$

to eliminate $T_{low}(t)$ and $\bar{T}(t)$; note that (4.27) is the solution of (4.28),

$$\frac{d\bar{T}(t)}{dt} = v(\bar{T}(t), t) \tag{4.28}$$

considering (4.20) and with initial condition $\bar{T}(0) = \bar{T}_o$. Surprisingly we obtain constant lower and upper thresholds for the normalised power curve $\Pi(t)$:

$$\frac{(T_{off} - T_{max})(\bar{T}_o - T_{min})}{(T_{max} - T_{min})(T_{off} - \bar{T}_o)} \leq \Pi(t) \leq \frac{(T_{max} - \bar{T}_o)(T_{max} - T_{on}) + (T_{off} - T_{max})(T_{max} - T_{min})}{(T_{max} - T_{min})(T_{off} - \bar{T}_o)}$$

(4.29)

If this restriction on the power levels and the temperature (energy) constraint (4.24) are both respected, it is guaranteed that each device can track the reference power profile $\Pi(t)$ in terms of its expected power consumption. This important result implies that there is no need to perform detailed device-level simulations to verify the actual feasibility of the response.

4.7 Device level problem and simulations

In Section 4.3 we discussed the framework of the advanced decentralized control proposed in [56]. The first stage, the ensemble control problem, has been introduced in Section 4.4, hence in this section, we briefly state the aim of the second stage of the control framework: the device control problem. This problem aims to state a device-level strategy that is compatible with $f(T,t)$ and $v(T,t)$ that result from the solution of the ensemble problem. Specifically, for the hybrid deterministic-stochastic controller presented in Section 4.3, this stage evaluates the switching rates r_{off}^{on} and r_{on}^{off} and the deterministic switching temperatures $T_{low}(t)$ and $T_{high}(t)$. After having defined a device-specific controller, single devices are switched on or off in accordance with the control strategy: a change in the appliance state always happens when the temperature hits the bounds $T_{low}(t)$ and $T_{high}(t)$, and stochastically according to r_{off}^{on} and r_{on}^{off} for intermediate temperatures. We refer the reader to Section II-D and IV-E of [56] for more insights about the solution of the device-level problem and a bottom up derivation of these results. In fact, for the remainder of this work, we will be focused only on the ensemble level problem. In doing so we take advantage of one of the key result of the ensemble level problem: there is no need to carry out device-level simulations to verify whether a certain response is feasible if the energy and power constraints are both verified. However, for completeness, we include the simulations of individual device actions in order to illustrate the results obtained in [56].

The default parameters for domestic refrigerators simulated are listed in Table 8 (Chapter 3). The devices are controlled by a common reference power curve $\Pi(t)$. By definition $\Pi(t)$= 1 represents steady state power consumption, and each response action has to start and end at this value (full recovery). For simplicity, the appliances are assumed to operate in 'fire and forget' mode: they track a pre-set power response $\Pi(t)$ which is triggered by the occurrence of a frequency event. This power profile is not adjusted dynamically in response to further changes in the power system

frequency, so the full trajectory of the control parameter $\beta(t)$ can be computed in advance. Figure 25 shows the *on-off* repetition for 20 single TCLs; the target power profile is shown at the bottom of the figure (dashed).

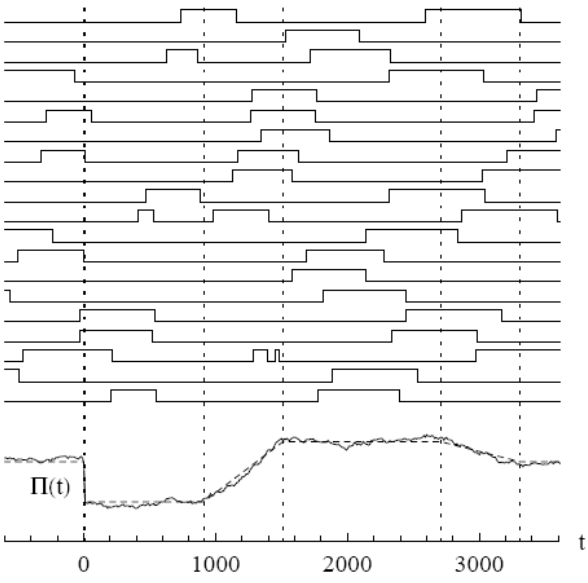


Figure 25 *On*/*off* status for 20 individual appliances shown alongside the reference power profile (bottom, dashed) and the empirical normalised power profile of $N = 1000$ aplliances (bottom, solid).

Note that after the initial load drop (linear reduction between $t = 0$ and $t = 10s$) the appliances tend to be in the *off* state, and thus the aggregate power consumption decreases. The opposite situation takes place during the payback phase, during which TCLs increment the switching to the *on* state. The empirical aggregate power profile generated by $N = 1000$ appliances (solid) is overlaid on the reference power profile for comparison.

4.7.1 Heterogeneous appliances

As we discussed in Section 4.2, the compatibility of a particular response profile should be verified by the demand response designer before broadcasting the desired response. The designer could resort to a database of appliance models to accomplish this request, but it is easy to imagine that creating and maintaining such a database is not a trivial task. Moreover, following this approach, a huge number of constraints must be checked. As an alternative we propose to use a simpler sufficient condition that include the aggregate capability of all devices into a single flexibility envelope. It follows that the power constraint (4.29) will be gearanteed for each TCL if and only if

$$\max_a(\Pi_{min}^a) \leq \Pi(t) \leq \min_a(\Pi_{max}^a) \tag{4.30}$$

Where $[\Pi_{min}^{a}, \Pi_{max}^{a}]$ is the range of accessible power levels for appliance a. In addition, the energy constraint (4.24) for each appliance is defined by the minimum sustained relative power level $\Pi_{sustained,min}^{a} = 1 - (T_{max}^{a} - \bar{T}_0^a)/(T_{off}^{a} - \bar{T}_0^a)$ and the thermal time constant τ^a. Equation (4.24) is satisfied for all appliances is, for all t,

$$\max_{a}\left(\Pi_{sustained,min}^{a}\right) \leq \frac{1}{\tau_{min}} \int_{-\infty}^{t} \Pi(t')\, e^{-\frac{t-t'}{\tau_{min}}} dt' \leq 1 \tag{4.31}$$

where

$$\tau_{min} = \min_{a} \tau^a \tag{4.32}$$

refers to the device with the lowest thermal time constant. Note that (4.31) is a sufficient but not necessary condition to ascertain the feasibility of the response, because the appliance with the highest value of $\Pi_{sustained,min}^{a}$ is not necessarily the one with the lowest value of τ^a. Following this approach, the response designer has to be aware of only four aggregate parameters to verify the feasibility of a solution. These four parameters can be obtained either through direct communication with the appliances or through intermediate demand side aggregators.

4.8 Impact of two different implementations of the advanced two-stage controller on system frequency control

So far, we summarised the key-characteristics and properties of the advanced two-stage controller developed in [56]. In this section, we proceed and compare two possible implementations of the controller for the network frequency support problem. The difference between them lies in the means through which the aggregate response is tailored. The first implementation retraces exactly the framework presented in the previous sections; the TCLs deliver a predetermined power response $\Pi(t)$ that is triggered by an initial frequency event. This response is not adjusted dynamically in response to changes in the network frequency; hence, the full trajectory of the control parameter $\beta(t)$ can be computed in advance. The second case results to be slightly more complicated as $\beta(t)$ is adjusted on the fly to track a dynamic power profile $\Pi(t)$ that is now function of the system frequency deviation $\Delta f(t)$ and its rate of change $\frac{d\Delta f(t)}{dt}$, individually or simultaneously.

Note that having such a reference power profile $\Pi\left(t, \Delta f(t), \frac{d\Delta f(t)}{dt}\right)$ does not to lead a to fundamentally different control strategy.

It is worth pointing out some changes in the notation; for the remainder of the chapter, if not differently specified, $f(t)$ will represent the network frequency [Hz] and not the probability density of the appliances anymore. We test the effectiveness of the two controller implementations on a basic but intuitive system model; this model mimics the dynamics of the future low carbon GB power system [12]. Computer simulations verify the increased ability of the network to integrate the expected renewable generation when TCLs are controlled with the proposed methods. Furthermore, the positive contribution of demand side response is studied not only on a short time scale interval (primary response [20]) but also on the longer time scale of secondary response [20].

4.8.1 The pre-programmed controller

The first application of the hybrid stochastic-threshold controller considers the tracking of pre-programmed power curve in response to frequency events. This response may be triggered by a sudden frequency drop, but, in the basic implementation considered, does not follow frequency after it commences (as discussed in Section 4.2). The ability to select an optimal response profile, in accordance with the amplitude of the power imbalance actually occurred, represents a further improvement, for instance, this would avoid an over-reaction and thus a positive frequency excursion when the net generation-demand imbalance is lower than the TCLs power reduction.

As illustrated in Chapter 2, in order to estimate the actual power imbalance, TCLs need to calculate the rate of change of frequency and recognize the level of system inertia H; in fact, by having awareness of these quantities, the amplitude of the power imbalance ΔP can be inferred by:

$$\Delta P = 2H\frac{df(t)}{dt} \tag{4.33}$$

Note that (4.33) holds only in the first seconds after the frequency event. Depending on the detected amplitude of ΔP, the appropriate pre-programmed response would be triggered. For the simulations carried out we assume for simplicity the perfect evaluation of ΔP; moreover we focus on the impact of the maximum power support from TCLs in order to face the maximum expected infeed generation loss. Hence, the TSO or the demand response aggregator has the freedom to design complex response $\Pi^{(1)}(t)$ according to system commercial/technical requirement. The only

hard limitations are enforced by constraints (4.23) and (4.29). Note that the superscript in $\Pi^{(1)}(t)$ denotes the first type controller, the pre-programmed control.

4.8.2 The frequency linear controller

The control architecture introduced in this chapter allows us to make a simple choice for tailoring of demand side response profiles. Hence, we propose

$$\Pi^{(2)} = \Pi\left(f(t), \frac{df(t)}{dt}\right) = K_1 \Delta f + K_2 \frac{d\Delta f}{dt} \tag{4.34}$$

This simple implementation has the desirable property that, after the demand response activation, the power consumption of the TCL population is always a linear function of the system frequency and its RoCoF. Initial temperature deadband controllers [28, 29] could not guarantee such a linear relation between power consumption and frequency. In fact, these versions only induce an initial power reduction that is proportional to the initial frequency drop; however, when frequency recovers, the long-term response shows a tendency for devices to synchronize their cooling cycles. Considering (4.34), four typologies of the linear implementation are proposed:

$$Case\ A: \quad K_2 = 0; \quad K_1 | \Pi(f_{\min}) = \Pi_{sust} \tag{4.35a}$$

$$Case\ B: \quad K_2 = 0; \quad K_1 | \Pi(f_{min}) = \Pi_{min} \tag{4.35b}$$

$$Case\ C: \quad K_1 = 0; \quad K_2 | \Pi(Rocof_{min}) = \Pi_{min} \tag{4.35c}$$

$$Case\ D: \quad K_1 | \Pi(f_{min}) = \Pi_{min}; \quad K_2 | \Pi(Rocof_{min}) = \Pi_{min} \tag{4.35d}$$

In Case A, when frequency achieves the minimum threshold, the reference power reaches $\Pi_{sust} = 1 - \left(\frac{T_{max} - \bar{T}_o}{T_{off} - \bar{T}_o}\right)$ (See equation (4.24)). By definition this power level can be sustained indefinitely without violating the energy constraint (4.23).This conservative assumption permits to ignore the dynamics of frequency recovery; if frequency is constantly maintained at f_{min}, the appliances will still be able to grant their support indefinitely. Case B augments the maximum demand support ($\Pi_{min} < \Pi_{sust}$); in particular, Π_{min} is the minimum accessible instantaneous limit in equation (4.29). Both cases A and B are insensitive to the RoCoF ($K_2 = 0$), whereas, Case C depends only on it. For this case during the very first instants of the frequency transient period the TCLs contribution is higher as the RoCoF is minimum (negative value); afterwards, the RoCoF becomes a positive quantity and thus the TCLs consumption is higher than the steady state level.

Under this implementation, the energy borrowed from TCLs is already paid back by the end of the primary response service. However, with large penetration of TCLs this choice could potentially cause positive frequency deviations during the time when RoCoF is a non-negligible positive quantity.

Finally, Case D tailors the power response depending on both frequency deviation and RoCoF; this version provides quicker and larger support as the control action is maximum since the first instants after the frequency drop until the time frequency reaches the nadir. The control action in this case is quite similar to the hybrid control described in Section 3.4.3. However these similarities only hold during the first instant of the network frequency transient period; afterwards the aggregate power response enabled by the hybrid controller loses the linear relation with frequency. Moreover, regarding cases B and D, the usual time scale of frequency restoration and the usual thermal capacity of refrigeration loads allow the TCLs to actually respect the energy constraint (4.23); however, the temperature evolution can be still monitored off-line using (4.13). For all the cases, the power response is never lower than Π_{min}.

4.8.3 Demand side response activation

An important aspect is the criterion for demand response activation; both the methods make use the grid frequency signal and its rate of change to trigger the TCLs frequency response. Figure 26 shows the activation scheme.

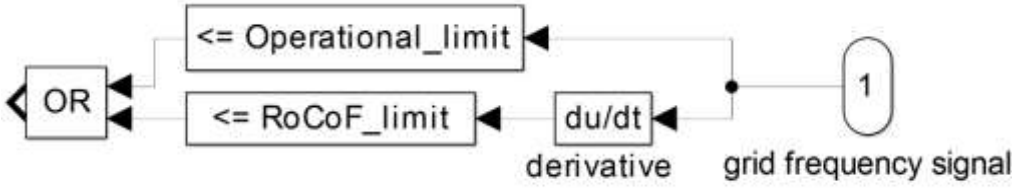


Figure 26 Demand side response activation

We consider two conditions that correspond to an emergency state for network frequency control. In the first case, frequency drops below the operational limit [20] ; in the second one instead, the RoCoF (negative quantity if frequency drops) is smaller than the minimum security threshold [20]. The occurrence of one of the two conditions or both at the same time (OR logic operator) triggers the demand response reduction.

4.8.4 Case Study

Two cases in accordance to the National Grid's 2020 Gone Green Scenario [12] are considered. For the high wind scenario 20 GW of wind power is supplying system demand; conversely, in the low wind case, the wind penetration is limited to 5 GW. For each scenario, we vary the system demand from 30 to 55 GW with a 5 GW step. Moreover, we assume that wind farms do not provide inertial support nor frequency response. The operational limit equals 49.8 Hz [20] , while the minimum RoCoF is set to -0.5 Hz/s. It is worth to point out that this value differs from current settings (-0.125 Hz/s [20]) and it is in agreement with NG suggestions for future low carbon scenarios [11].

The standard linear frequency response model presented in [58] is extended to integrate the demand response support and the reserve provided by generators (Figure 27). The model considers the contribution of primary and secondary control. In addition, the provision of secondary response from generators is maintained for 30 minutes after the failure [20] ; afterwards, the power provided is reduced linearly (in 15 minutes) and replaced by reserve through additional units.

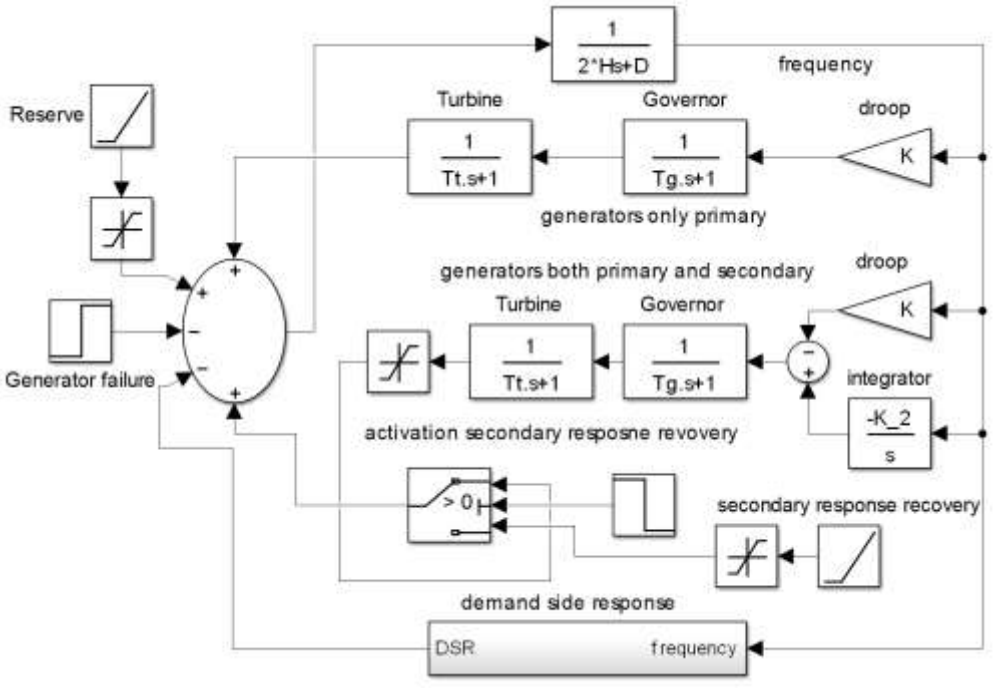


Figure 27 Linear system frequency response model

The parameters of this model are chosen to qualitatively match the frequency dynamic behaviour of the GB system; In particular, the time constants for the governor and the turbine are $t_g = 1s$ and $t_t = 10s$. The constant of inertia for the synchronous generators is 4.5s, while the

values for the gains are $K = 3.33$ and $K_2 = 0.025$. The load damping effect $D = 1$ is in units of MW/Hz and thus it has to be multiplied for the total system demand and divided by the nominal frequency. A step function simulates the sudden generator failure; the amplitude of the step is 1.8 GW: this value represent the maximum size of a generation loss for the future GB system [12].

For this set of simulations common domestic fridge-freezers units are taken into account; each individual appliance is described by a first order thermal model (see Section 3.3.2). The parameters for the related appliances are listed in Table 8 (Chapter 3). The number of refrigerators for the GB context is assumed to be 45 million and thus the steady state power of the TCLs population is 1.77 GW. The TCLs' power response with the frequency linear controller is exclusively driven by the frequency and RoCoF evolution. With the alternative method, the power profile is designed *off line*. However, since the optimal allocation of frequency response services is also driven by technical and economic issues, at this step we do not aim to determine the optimal shape of the reference response profile $\Pi^{(1)}(t)$; hence, we take as a reference target curve the blue solid profile in Figure 28.

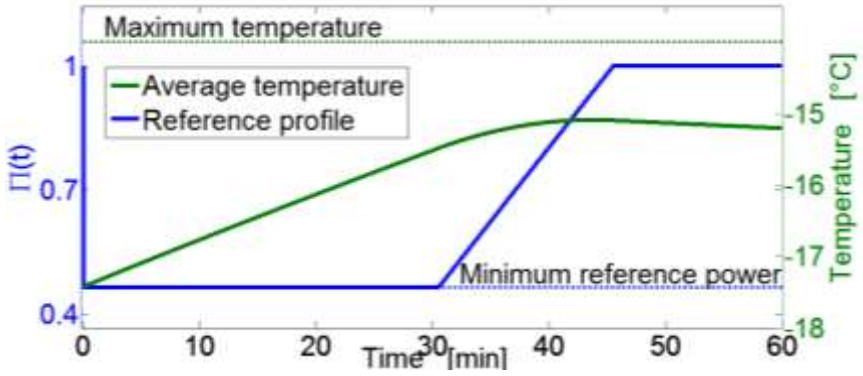


Figure 28 Reference power profile (blue solid) and consequent average temperature evolution (green solid).

Individual appliances can follow this reference profile as it respects the power limits (4.29) and because the average temperature (green solid profile) satisfies condition (4.23). Note that the selected profile $\Pi^{(1)}(t)$ maximizes the TCLs' support for primary and secondary response. The frequency drop triggers a demand reduction that achieves the lower boundary of equation (4.29) after one second. This reduction is maintained for 30s (primary response) and for further 30 minutes (secondary response). Afterwards, the profile recovers the pre-fault value through a ramp (15 minutes). As we are not optimizing the recovery of the average temperature to the steady state value, we do not include the payback phase in the reference power profile. However, the controller

would able to boost the power consumption in order to speed up the recovery of the average temperature.

4.8.5 Results

The effectiveness of the control methods proposed is illustrated in Figure 29 for both the low (29A) and high (29B) wind scenarios. Without implementing any demand control scheme, the frequency nadirs (violet) drop below the security limit (grey) in most of the cases (especially in the high wind scenario). The results do not improve significantly making use of the versions A (yellow) and C (green) of the frequency linear controller. In fact, although the former permits only a poor maximum power reduction $\Pi(f_{\min}) = \Pi_{sust}$, the latter, following the RoCoF dynamics, contributes mostly in the first instants whereas the support is nil at frequency's nadir (RoCoF=0).

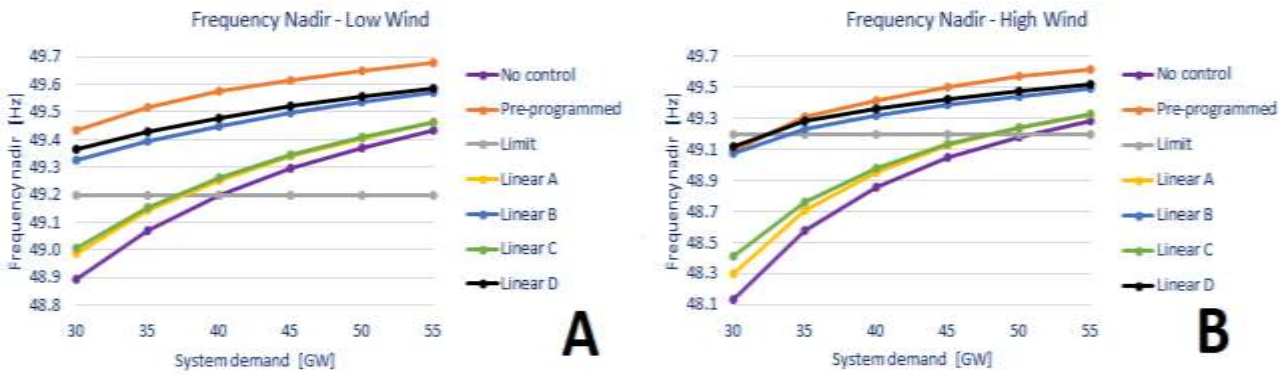


Figure 29 Frequency nadir for low wind (A) and high (B) wind penetration

Conversely, the linear controller (Case B, blue and D, black) and the pre-programmed method (orange) provide overall beneficial results. In particular, the frequency nadirs are always above the security limit; this implies that the total wind integrated can actually supply the system demand without being curtailed. Only in one severe scenario (30 GW demand and 20 GW of wind), the TCLs support cannot limit the frequency drop; a deeper penetration of controllable TCLs (e.g. commercial refrigeration) will allows for a full wind integration. Let us now compare with more details the three most effective solutions. Case D performs slightly better than Case B; the power reduction is quicker as it exploits the RoCoF evolution that is maximum just after the frequency drop.

On the other hand, the pre-programmed controller obtains, in most of the cases, the lowest frequency nadirs. For few severe cases of the high wind scenario, the performance of pre-programmed controller and the linear (Case D) are equivalent; frequency falls vary fast (high RoCoF),

therefore the initial power reduction with the linear controller retraces the fast reduction of the pre-programmed power profile.

For the remainder of the section, we consider the scenario with 45 GW of demand, 20 GW of wind and the remaining difference supplied by conventional generators. The complete frequency evolution and restoration of the steady state condition is shown in Figure 30A; the zoom on the first instants after the generation loss helps to highlight the effectiveness of demand side response (Figure 30B). We now select the most effective control strategies, the pre-programmed (red) and the Case D of the frequency linear controller (black); in fact, Cases A-C are simply reduced version of Case D. With both the strategies, the frequency nadir is far away from the security limit of 49.2 Hz.

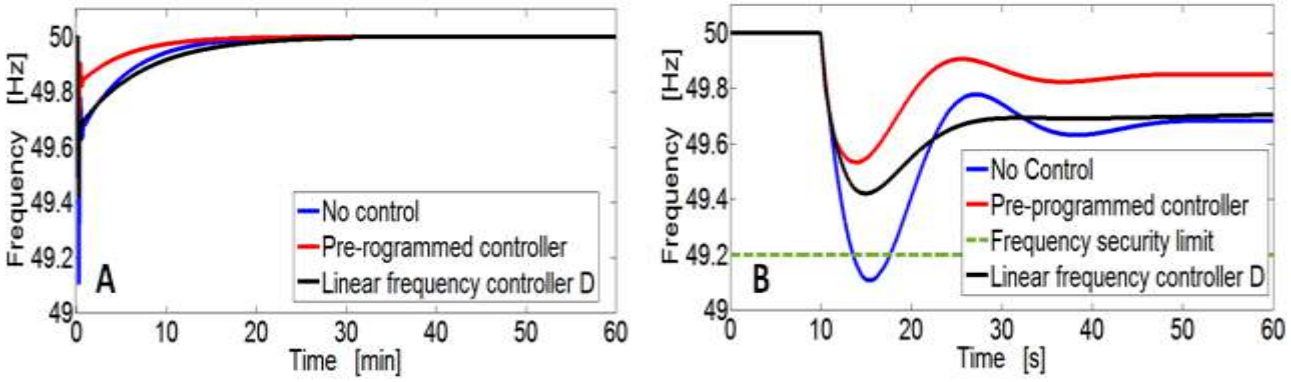


Figure 30 (A) Frequency evolution and restoration with: pre-programmed controller (red), frequency linear controller D (black) and without demand support (blue). The zoom in (B) shows the fast transient dynamics during primary response.

Moreover, the frequency evolution with the linear controller is more damped. This effect is consequence of the positive value of the RoCoF after that frequency achieves the nadir. In fact, during this phase, the second term of (4.34), $K_2 \frac{d\Delta f}{dt}$, acts in an opposite way compared to the first term, decreasing the overall power reduction. Furthermore, this opposite behaviour terminates quickly as the RoCoF tends to zero after the primary response provision, while the frequency deviation is a constant negative quantity. We now increase the time window to assess the benefits of demand side response in the secondary response/reserve time interval. The provision of secondary response with the pre-programmed strategy makes the restoration of frequency around the nominal value is much quicker. Conversely, the linear controller imposes a slightly slower frequency recovery compared to the reference case; again, this is due to a positive and tiny value of the RoCoF during this phase. This effect is now smoother than during the primary response as the amplitude of the RoCoF is smaller.

The impact of demand side support on generators' response and reserve is presented in Figure 31; these results demonstrate the ability of demand response to replace fast frequency services provided by part-loaded generators or fast but high-cost units. The primary response provided by generators reduces drastically thanks to TCLs' contribution with both controllers (red and black solid) compared to the reference case. The amount of power supplied by generators in charge of primary and secondary response is significantly reduced with the pre-programmed controller (dash red); in addition, the power ramp rate to provide secondary response decreases, facilitating the replacement of fast but expensive and pollutant generators with other units. The frequency linear controller cannot guarantee similar benefits as the TCLs participation in the secondary response is poor as the TCLs response follows the frequency that is now already close to 50 Hz. The reserve profile does not change; however, the generators that now do not provide secondary response anymore can facilitate the fulfilment of the reserve requirement.

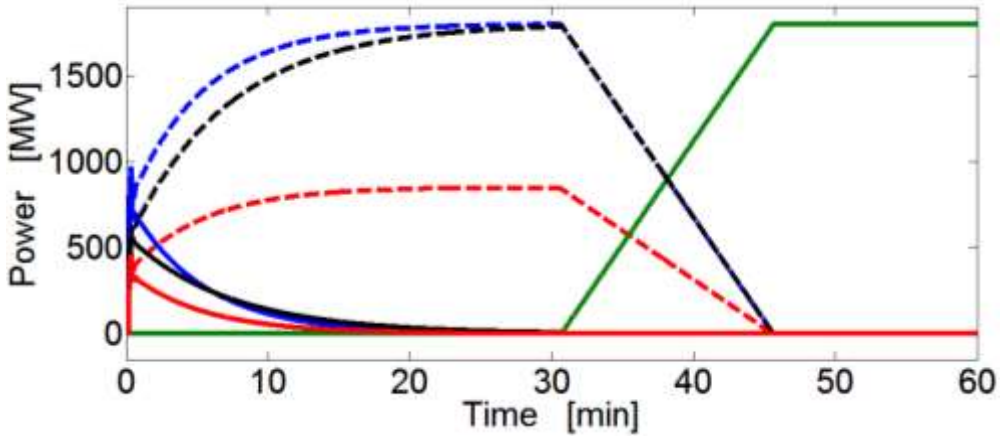


Figure 31 Power supplied by generators that provide only primary response with: pre-programmed controller (red solid), frequency linear controller (black solid) and without demand support (blue solid).Power provided by generators in charge of primary and secondary response with: pre-programmed controller (red dashed), frequency linear controller (black dashed) and without demand support (blue dashed). Reserve power provision supplied by further units (green solid).

4.9 Conclusions

In this chapter we have discussed the framework and the main outcomes of a novel control architecture for controlling the aggregate power consumption of thermostatically controlled loads. We found that this controller is able to simultaneously address four criteria that define the optimal use of demand response capability; other methods in the literature cannot accomplish all of them at the same time. The underlying idea of this methodology is to control each appliance

independently in such a way that its expected power consumption tracks (in expectation) a target relative power consumption profile Π(t). Considering the realistic case of large number of TCLs, the statistical fluctuations will average out and thus the population power consumption targets a well-defined value $P_0^{total}\Pi$(t). However, the choice of the target profile $\Pi(t)$ is limited only by the physical characteristics of the devices and the need to not significantly impact the primary function of these devices[4]. Moreover, the control framework can be implemented in a fully decentralised fashion and it is able to guarantee that no single TCL exceeds the temperature range $[T_{min}, T_{max}]$.

The peculiarity of this method lies in the approach to the control problem; in contrast with alternative solutions in the literature, the control approach is inverted by making use of a collective property, the net heating rate $v(T, t)$, as the control parameter. Only in a second phase, occasional stochastic switching rates and the threshold temperatures are then derived from $v(T, t)$ to physically manipulate the device operation, instead of the other way around. Hence it is possible to separate the control problem into two stages: a first stage (ensemble problem) in which the heating rate profile is computed to permit the tracking of the ensemble power consumption Π(t), and a second stage (device-level problem) in which this solution is translated to a controller for an individual appliance.

Moreover, this high-level analysis has been applied to the case where TCLs' temperatures were described by first order ODEs. A general result was inferred for such a basic but effective model; it is straightforward to compute the evolution of the ensemble average temperature from the aggregate power consumption - regardless of the specifics of the controller. Further outcomes also proved that a so-called 'payback' period of increased power consumption is not strictly required after a power reduction: if TCLs are operated at the steady-state power level they will eventually recover the steady-state temperature asymptotically. However, in practical applications it will often be desirable to include a payback period in order to speed up the return of the average temperature to its steady state value In fact, the controller considered would allow the system operator to optimally allocate the payback profile in accordance with the system's characteristics. The economic value of allowing a payback period to accelerate the temperature/energy recovery after the provision of medium-term response services is quantified in Chapter 7.

[4] The optimal reference profile is also driven by economic aspects; this issue will be analysed in the next chapters.

Furthermore, we have introduced a specific controller design where the heating rate $v(T,t)$ results to be a linear function of the temperature T. For this specific case, a comparison with the generic limitations for linear thermal models highlights that the structure of designed linear controller is optimal with respect of the energy provision for load reduction. Moreover, as explained in section 4.7.1, the system operator can easily drastically reduce the burden to guarantee the feasibility of the response of a heterogeneous cluster of device by using a 'lowest common denominator' envelope model for their collective flexibility. Again, in the case of the linear controller, this flexibility envelope is characterised by only four parameters. Together with the average aggregate power consumption P_0 these can be used to design a feasible demand response pattern.

At the end of this chapter, we evaluated the impact of two different implementations of the advanced decentralized controller on the system frequency control. Both a frequency linear controller and a pre-programmed controller suit the overall high-level control framework whereas the main difference lies in the choice of the reference power profile Π(t); in the first case, this profile is always proportional to the frequency evolution while in the second case it can be arbitrary. Both strategies provide promising results towards the integration of RES in a future power system. In particular, we proposed further different solutions for the frequency-linear controller.

A poor support to the network is revealed if the maximum power reduction is limited to Π_{sust} (Case A) or it is only proportional to the RoCoF (Case C), while those controllers that provide a larger power reduction Π_{min} and are proportional to the frequency deviation (Case B) and also to its rate of change (Case D) to be more performing in the system frequency support. Finally, the pre-programmed controller offers a large and more flexible support to the frequency control; the TCL full support can be extended to the secondary control time window, enabling further support from the devices. Moreover, with this method, the system operator or demand response aggregator has the freedom to design complex responses in accordance with the technical and commercial requirements of the network. The better performances offered by the pre-programmed controller require a slightly more sophisticated control infrastructure as TCLs need to estimate the level of the system inertia to provide the correct power reduction, avoiding useless over-reactions. The approximate system inertia may be predicted with reasonable accuracy (see equation (4.33)), thus this information could be broadcast by means of smart meters. The communication delays involved in this channel would not be critical. On the other side, the frequency linear controller does not

require this information and would be able to work in a *fully autonomous* way. Finally the pre-programmed controller would be preferable as it can automatically implement longer-term power profiles (on top of emergency power reductions for ancillary services) that exploit the ability of TCLs to realize energy arbitrage (Chapter 5).

4.9.1 Extensions and links with the next chapters

The linear controller chosen may serve as a basis for the development of more sophisticated controllers; we disclose few extensions to the controller framework. Here we limit to recall the extensions that have been considered and applied in the next chapters of this thesis. The controller may be expanded (see next chapter) to eventually enable the absorption of excess energy from the grid (high-frequency services [76]). In addition, as the control framework in presented is be able to precisely control the aggregate consumption across a range of time scales, we will investigate the applicability of this property for the provision of energy arbitrage services, where TCLs are used to relax constraints on the dispatch of generators (Chapters 5, 6 and 7). Thermostatic loads can reduce their power consumption at times when the on-line generators are very expensive or carbon-intensive, and delay their power consumption to adjacent time intervals. Of course, when the periods characterized by low and high prices are many hours apart (e.g. night and day), the thermal storage capacity of the TCLs would not be able to trigger significant load transfer between these time windows. If that is the case, the TCLs' flexibility may still be exploited to alleviate generators' ramping constraints between such periods.

The ability to track an arbitrary relative power profile $\Pi(t)$ allows to grant both frequency and energy arbitrage services simultaneously, provided that such a simultaneous provision respect the physical constraints of the appliances. In such a scenario, in normal operation TCLs would follow a certain power profile to make energy arbitrage and provide scheduled frequency services (switching on a different power profile) only if a frequency drop is observed. The appliance's power response model would contain both components and should be designed in such a way that the appliance always maintains sufficient energy buffer to deliver the scheduled frequency services if required. In particular, Chapter 5 explores the benefits of this ability to select optimal portfolios of energy arbitrage and frequency response services in accordance with the TCLs physical constraints.

Finally, Chapters 6 and 7 attempt to determine the optimal choice of $\Pi(t)$. This optimal choice is the result of an interplay between technical and economical aims, and the limits of the

appliances and their controller. Determining the optimal power profile $\Pi(t)$ is the necessary condition to assess the value of demand response services.

Chapter 5: *The Leaky Storage Unit for Optimal Multi-service Allocation of Thermostatic Loads*

Abstract

Thermostatically controlled loads offer significant potential for short term demand response. This intrinsic flexibility can be used to provide a range of frequency regulation and reserve services or to carry out energy arbitrage. This chapter addresses the simultaneous provision and optimal allocation of multiple services that maximises the revenues for a portfolio of TCLs. Starting with physical device models and quality of service constraints an envelope model is developed for the flexibility of a heterogeneous aggregation of TCLs. This model is used to illustrate the diversity in the capabilities of eight representative classes of cold appliances. The flexibility model has an equivalent representation as a *leaky storage unit* with associated constraints; this model is embedded in a linear optimisation model that determines the optimal multi-service portfolio for a cluster of TCLs. The service allocation and resulting profits reflect the differences in physical capabilities between the appliance classes.

5.1 Introduction

In the previous chapters, we focused on the ability for TCLs to provide frequency services in order to support the network frequency control. The fast-responding flexibility of these devices releases conventional generators from providing response services so they no longer need to be operated part-loaded; moreover, such a demand side support helps to contain the frequency deviations below security thresholds even in presence of a large share of RES that do not participate to the system's inertial-frequency response. We also investigated the possibility for TCLs to supply long-term response services; this ability reduced the need for operating high cost plant eventually characterized by high emissions.

However, the potential applications and benefits for TCLs are not limited to the frequency control; the intrinsic flexibility of TCLs may be used for other applications. For example, time-dependent tariffs may encourage TCLs to perform energy arbitrage, thus consuming more energy when electricity prices are low and less when high [48, 34]. In sufficient quantities, the actions of TCLs may decrease peak load and thus reduce the need for high-cost peaking plants. Even when limits to thermal inertia restrict the potential for energy arbitrage, the TCLs' flexibility may still be used to alleviate ramping constraints. Furthermore, a flexible TCL power consumption may be exploited to alleviate transmission system congestions. Hence, there is a clear value proposition for flexible TCLs, but two significant technical challenges must be overcome to make optimal use of this flexibility. The first challenge is how to accurately control the aggregate power consumption of a collection of TCLs without violating device-specific temperature constraints. The accurate control of these devices is not straightforward [49] mainly due to their peculiar normal operation mode which implements only two power states (maximum power consumption when *on* and negligible consumption when *off*).

The first challenge was faced in Chapter 4 where we introduced a novel TCL control strategy that is able to provide accurate decentralised control of heterogeneous TCLs whilst respecting strict temperature bounds for each appliance. This control architecture overcomes the limitations associated with other controllers; in particular, its fully decentralised approach enhances the class of controllers that rely on a centralised control scheme to provide the desired level of controllability [41, 42, 33, 34]. Furthermore, the control framework perfects the distributed stochastic controller

[31] that provides frequency response services, but it is unable to simultaneously guarantee short-term load control and strict observance of per-device constraints (mainly linked to the temperatures).

The second challenge is to co-ordinate the provision of multiple interdependent system services, with the overall objective to support profit-maximising operation of TCLs. A realistic implementation of such demand response schemes involves the matching of physical characteristics of TCLs with the services they can provide and selecting optimal portfolios of services that can be to be provided simultaneously. Previous research on TCLs has largely focused on their ability to provide individual services [29, 34, 77].

This chapter addresses the second of these challenges and explores the price-driven optimal allocation of energy arbitrage and frequency services for a variety of cooling appliances. A significant contribution is the *leaky storage model*, which describes the aggregate flexibility of a heterogeneous population of TCLs as a single storage unit with a leakage/evaporation term. This intuitive representation allows for an easy embedding in an optimal service allocation framework. The leaky storage model incorporates the physical thermal models of heterogeneous TCLs and their operational temperature limits. The model is exact for large populations of TCLs that can be described by constant first order models and – in combination with a suitable controller – it guarantees that scheduled responses can be delivered by the least flexible appliance. We analyse the eight representative classes of cooling appliances, introduced in Chapter 3; the large differences in their physical parameters detected at the modelling stage result in different allocations of services to each class.

The chapter is organised as follows; Section 5.2 summarises key results regarding the modelling and controllability of a heterogeneous population of TCLs. Section 5.3 deals with the fundamental ability of these devices to provide various services to the grid; the relative strengths and weaknesses of each TCL class are demonstrated by making use of effective aggregate model parameters. The representation of a heterogeneous cluster of TCLs as a leaky storage unit is presented in Section 5.4. Moreover, Section 5.5 describes the linear optimization model for the allocation of energy arbitrage and frequency response services. The optimal solution of the problem is presented in section 5.6. Finally, Section 5.7 concludes the chapter and it is followed by the Appendix.

5.2 Aggregate thermostatically controlled load modelling

This section develops an aggregate model for flexible thermostatically controlled loads, using steps analogous to those in Section 4.6. The derivation presented here focuses on system-level properties.

5.2.1 Temperature and power

In the most common implementation a thermostatically controlled appliance a switches between two states: passive cooling/heating (no power consumed) and active heating/cooling (power consumption P_{on}). This system is commonly modelled by a first order temperature model as described with more details in Section 3.3.2. Hence, we recall the first order model using a compact formulation of equation (3.7); the following differential equation describes the evolution of the appliance temperature $T^a(t)$:

$$\frac{dT^a(t)}{dt} = -\frac{1}{\tau^a}\left[T^a(t) - {T_{off}}^a + s^a(t)\cdot\left({T_{off}}^a - {T_{on}}^a\right)\right] \tag{5.1}$$

here $s^a(t)$ is the binary cooling state of appliance a, 1 (*on*) or 0 (*off*), τ^a the thermal time constant, ${T_{off}}^a$ the ambient temperature and ${T_{on}}^a$ the asymptotic cooling/heating temperature, which incorporates a physical model of heat exchange. Considering a population of N TCLs, equation (5.2) follows from the linearity of (5.1)

$$\frac{1}{N}\sum_a \frac{dT^a(t)}{dt} = -\frac{1}{N}\sum_a \left[\frac{T^a(t) - {T_{off}}^a + s^a(t)\cdot\left({T_{off}}^a - {T_{on}}^a\right)}{\tau^a}\right] \tag{5.2}$$

This property indicates that the mean temperature $\frac{1}{N}\sum_a T^a(t)$ of a large population of TCLs evolves as equation (5.3). In effect, taking into account a population of TCLs with identical parameters τ, T_{on}, T_{off} (these parameters no longer need the superscript a) we obtain

$$\frac{d\bar{T}(t)}{dt} = -\frac{1}{\tau}\left[\bar{T}(t) - T_{off} + \pi(t)\cdot\left(T_{off} - T_{on}\right)\right] \tag{5.3}$$

Where $\bar{T}(t) = \frac{1}{N}\sum_a T^a(t)$ and $\pi(t) = \frac{1}{N}\sum_a s^a(t) \in [0,1]$ is the fraction of appliances in the *on* state. Therefore, the average temperature of a population effectively behaves as the hypothetical temperature of a single large appliance with a variable cooling rate. Note also that a population of identical appliances with unbiased random initial conditions may be interpreted as a

probability distribution for the state of a single appliance with unknown initial conditions. The average properties of such a population are therefore representative of expectation values for a single appliance, which implies that $\pi(t)$ is proportional to the expected power consumption $P^a(t)$ of a single appliance:

$$\pi(t) = E[P^a(t)]/P_{on} \tag{5.4}$$

From this perspective equation (5.3) effectively relates the expected power consumption $E[P^a(t)]$ and temperature $E[T^a(t)] = \bar{T}(t)$ of an appliance, and any controller that modulates the expected power consumption changes the expected temperature as a consequence. In addition, equation (5.3) determines the relation between the steady state probability of being in the *on* state π_o (the duty cycle) and the steady state expected temperature T_0:

$$\pi_o = \frac{T_{off} - T_0}{T_{off} - T_{on}} \tag{5.5}$$

The following coordinate transformation is applied in order to streamline the further analysis:

$$\Pi(t) = \frac{\pi(t)}{\pi_0} \tag{5.6a}$$

$$\sigma(t) = \frac{T_{off} - E[T^a(t)]}{T_{off} - T_o} \tag{5.6b}$$

where $\Pi(t)$ is the relative power level and $\sigma(t)$ is a dimensionless measure of stored energy (explored in later sections). Expressed in these variables, equation (5.3) becomes

$$\frac{d\sigma(t)}{dt} = -\frac{1}{\tau}[\sigma(t) - \Pi(t)], \tag{5.7}$$

with the steady state solution $\sigma_0 = 1$ and $\Pi_0 = 1$. Note that the analysis above is applicable to any first order TCL.

5.2.2 Controller and quality of the service constraint

The primary function of a TCL is to maintain a compartment close to a desired temperature T_0. Fluctuations around this target must be limited, resulting in bounds on $\sigma(t)$. In addition, both $\sigma(t)$ and $\Pi(t)$ may be further constrained by the actual TCL controller implementation. In this chapter we assume that the TCL actions are coordinated by the two-stage stochastic-threshold

controller proposed in Chapter 4. This is a decentralised controller that takes a control signal $\Pi(t)$ and controls the expected power consumption of each individual appliance in accordance with equation (5.4), so that the aggregate power consumption of $N \gg 1$ identical appliances is

$$P(t) \cong N\pi_0 P_{on} \Pi(t) \tag{5.8}$$

We assume that an aggregator represents the TCLs on relevant markets and broadcasts the contracted services to the appliances ahead of time. The controller guarantees that no appliance ever exceeds this acceptable temperature interval as a result of a control action. This property implies that the quality of service condition

$$T_{min} \leq T^a(t) \leq T_{max} \tag{5.9}$$

for the appliance temperature $T^a(t)$ is equivalent to the (usually weaker) condition

$$T_{min} \leq E[T^a(t)] \leq T_{max} \tag{5.10}$$

In terms of the dimensionless parameter $\sigma(t)$ defined in equation (5.6b) this results in the constraint

$$\sigma_{min} \leq \sigma(t) \leq \sigma_{max} \tag{5.11}$$

In combination with equation (5.7) this equation limits the duration and extent of changes in the relative power level $\Pi(t)$. The controller also imposes further constraints on the instantaneous relative power consumption $\Pi(t)$. Note that the linear controller proposed in Chapter 4 was constructed for the purpose of power reduction and subsequent recovery alone. For the purpose of the present study we implement the extensions suggested in Section 4.9.1; in practice, we make use of a generalised controller that links this power-reducing controller to its power-increasing counterpart that scales the temperature distribution with respect to T_{min}, without loss of generality. The resulting controller is able to make use of the full range of equation (5.11) and switching between the two regimes takes place when $\bar{T}(t) = T_0$ or $\sigma(t) = 1$. Let us denote with Π_{min}^{down} and Π_{max}^{down} the instantaneous relative power limits when the controller is in the power-reducing mode (See equation (4.29) for the mathematical expression). Retracing the steps developed in Chapter 4, it can be verified that the resulting limits for the power-increasing counterpart (with $v(T,t) = \frac{\beta(t)}{\tau}(T - T_{min})$, $T_{low}(t) = T_{min}$ and $T_{high}(t)$ variable) are

$$\Pi_{min}^{up} = \left[\frac{\left(T_{off} - T_{min}\right)\left(\bar{T}_o - T_{max}\right)}{\left(T_{min} - T_{max}\right)\left(T_{off} - \bar{T}_0\right)}\right] \tag{5.12a}$$

$$\begin{aligned}\Pi_{max}^{up} = \big[T_{max}\left(T_{min} - T_{off}\right) - \bar{T}_o\left(T_{min} - T_{on}\right) \\ + T_{min}\left(T_{off} - T_{on}\right)\big] / \big[\left(T_{max} - T_{min}\right)\left(\bar{T}_0 - T_{off}\right)\big]\end{aligned} \tag{5.12b}$$

as the combined version of the stochastic-threshold controller would switch mode every time results that $\bar{T}(t) = T_0$ or $\sigma(t) = 1$, we take the conservative assumption for $\Pi(t)$ so that,

$$\max\left(\Pi_{min}^{up}, \Pi_{min}^{down}\right) = \Pi_{min} \leq \Pi(t) \leq \Pi_{max} = \min\left(\Pi_{max}^{up}, \Pi_{max}^{down}\right) \tag{5.12c}$$

5.2.3 Aggregate power control of heterogeneous appliances

The preceding sections show that – using the right controller – it is possible to modulate the aggregate power consumption $P(t)$ of a collection of identical TCLs using a control signal $\Pi(t) = P(t)/P_0$, where P_0 is the steady state power consumption. This flexibility is constrained by equation (5.12c) (instantaneous power) and the combination of equations (5.7) and (5.11) (integrated power, i.e. energy). However in practice a large group of TCLs will never be homogenous, so one must consider the restrictions imposed by heterogeneous appliances. Using the decentralised control strategy introduced in Chapter 4, each appliance can individually target the desired relative power curve $\Pi(t)$, so the desired response is guaranteed as long as each appliance can satisfy constraints (5.11) and (5.12c) according to its appliance-specific model parameters. This can in turn be guaranteed by enforcing the following envelope constraints:

$$\widehat{\Pi}_{min} \leq \Pi(t) \leq \widehat{\Pi}_{max} \tag{5.13a}$$

$$\hat{\sigma}_{min} \leq \sigma(t) \leq \hat{\sigma}_{max} \tag{5.13b}$$

$$\Pi(t) = P(t)/\hat{P}_0 \tag{5.13c}$$

where

$$\frac{d\sigma(t)}{dt} = -\frac{1}{\hat{\tau}}\left(\sigma(t) - \Pi(t)\right) \tag{5.14a}$$

$$\widehat{\Pi}_{min} = \max_a \Pi_{min}^a \qquad \widehat{\Pi}_{max} = \min_a \Pi_{max}^a \tag{5.14b}$$

$$\widehat{\sigma}_{min} = \max_a \sigma_{min}^a \qquad \widehat{\sigma}_{max} = \min_a \sigma_{max}^a \tag{5.14c}$$

$$\hat{\tau} = \min_{a} \tau^{a} \qquad \hat{P}_0 = \sum_{a} P_0^{a} \tag{5.14d}$$

and the superscript a is used for appliance-specific parameters. This representation reverts to that in equations (5.7), (5.11) and (5.12) when all appliances are identical. It is worth pointing out that the envelope parameters (5.14b - 5.14d) do not necessarily represent a real (or reachable) device as the device with the lowest thermal time constant $\hat{\tau}$ does not necessarily have to be characterised by the maximum $\widehat{\sigma}_{min}$ and minimum $\widehat{\sigma}_{max}$ at the same time.

5.3 Physical models and flexibility

The results in the previous section establish a link between physical device models and the aggregate flexibility of TCLs, expressed through the dimensionless parameters $\Pi(t)$ and $\sigma(t)$ and the thermal relaxation time τ. This section builds on those results to gain insight into the fundamental ability of TCLs to provide various services to the grid. We consider eight distinct types of refrigeration units across the domestic and commercial sectors. Refrigeration units represent a substantial base load and are connected to the grid at nearly all times, making them suitable for providing frequency services around the clock. For the domestic sector we distinguish refrigerators, freezers and fridge-freezers (see Section 3.3.1), while within the commercial area (see Section 3.3.2) we take into account bottle coolers (with glass doors), upright refrigerators, upright freezers, and two different multideck refrigerators (open retail units).

Tables 7 and 8 (in Chapter 3) summarize the characteristics of the representative appliances. Afterwards these representative parameters are independently varied by +/-15% to generate a heterogeneous set of appliances for that class. Aggregate envelope parameters for each class are computed using equations (5.13)-(5.14). Note that instead of T_{max} and T_{min}, we vary $(T_{max} + T_{min})/2$ and $(T_{max} - T_{min})$ by +/-15% to simulate variability in the set-point temperature and deadband width. The entire set of parameters for the heterogeneous cluster of TCLs is listed in Table 9. We consider the suitability of each appliance class for energy arbitrage and frequency support services. Three different frequency services are modelled in this chapter, according to the National Grid practice: primary response, high frequency response, secondary and flexible response [78].

Table 9 Thermostatic loads characteristics

TCL class	Heterogeneous cluster parameters					
	$\hat{\tau}$ [h]	$\hat{\mathbf{P}}_0/N$ [W/unit]	$\hat{\mathbf{\Pi}}_{min}$	$\hat{\mathbf{\Pi}}_{max}$	$\hat{\sigma}_{min}$	$\hat{\sigma}_{max}$
Domestic refrigerator	1.7	16.8	0.59	1.99	0.89	1.11
Domestic freezer	4.4	24.7	0.54	2.03	0.94	1.06
Domestic fridge/freezer	4.3	39.5	0.55	2.21	0.93	1.07
Bottle cooler	4.7	139.6	0.66	1.08	0.88	1.10
Upright fridge	12.9	60.0	0.54	1.51	0.95	1.05
Upright freezer	16.3	100.4	0.53	1.48	0.97	1.03
Multideck 1	0.5	524.0	0.62	1.11	0.91	1.08
Multideck 2	0.8	503.0	0.60	1.19	0.91	1.09

The primary and high-frequency response services are *short-term services*, requiring a constant decrease or increase in power consumption for up to 30 seconds. The secondary and flexible response services are *medium term* services. Both require a constant reduction in power consumption for up to 30 minutes, but they differ in their contracted hours.

5.3.1 Short term services

The ability to reduce or increase power consumption at short notice in order to provide short term services such as primary/high frequency response is directly linked to the constraints on $\Pi(t)$ (equation (5.13a)). Energy constraints do not play an appreciable role due to the short duration of these services. Figure 32 (shaded regions) shows the accessible power ranges for each TCL class, based on equation (5.13a).

Thin vertical lines show the hypothetical limits for homogeneous clusters of the reference appliances; as expected, the heterogeneous clusters have less flexibility. The domestic units have low steady state duty cycles and an ability to significantly increase their short-term power consumption. In contrast, bottle coolers and multidecks are only able to provide limited high frequency response.

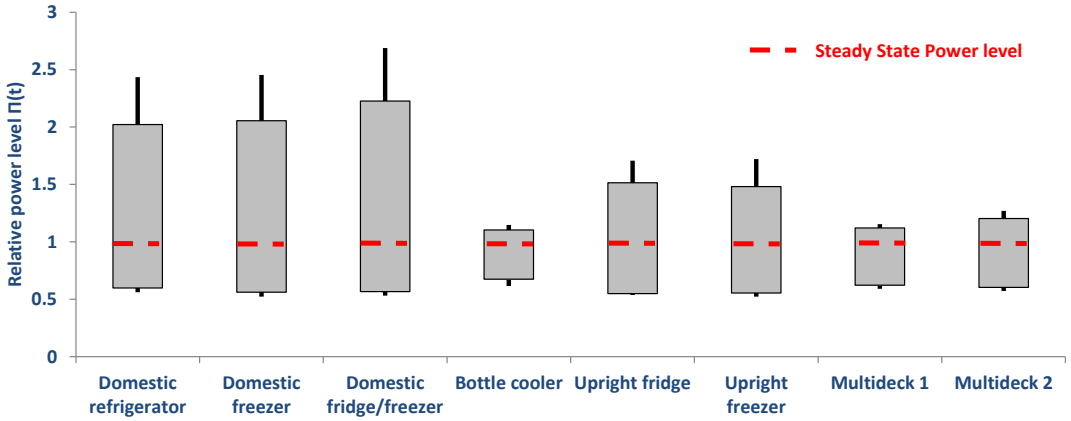


Figure 32 Ability to reduce/increase the power consumption at short notice: the red markers represent the steady state power level. Upper/lower limits of the shaded regions are the possible downwards/upwards power excursions. Vertical lines represent the same limits for a homogeneous cluster.

5.3.2 Medium term services

Under the secondary frequency response and flexible frequency response services a population of TCLs is requested to decrease its overall power consumption by $\Delta\Pi$ and to sustain this for 30 minutes [78]. To approximate this scenario we solve (5.14a) with the initial condition $\sigma(0) = 1$ (steady state) and $\Pi(t \geq 0) = 1 - \Delta\Pi$ and find

$$\sigma(t) = 1 - \Delta\Pi \cdot \left(1 - e^{-t/\hat{\tau}}\right), \quad \text{for } t \geq 0 \tag{5.15}$$

with the asymptotic limit $\sigma(t \to \infty) = 1 - \Delta\Pi$. When $1 - \Delta\Pi$ is less than $\hat{\sigma}_{min}$ such a demand reduction response can be sustained only for a limited time until $\sigma(t^-) = \hat{\sigma}_{min}$. This maximum response duration is given by

$$t^- = \hat{\tau} \log\left(\frac{1}{1 - (1 - \hat{\sigma}_{min})/\Delta\Pi}\right) \tag{5.16}$$

The aptitude of different classes of TCLs for providing long term services is illustrated in Figure 33. The dashed curves are obtained by solving equation (5.16) for constant $\Delta\Pi$, imposing $t^- = 30$ minutes. The iso-power-reduction curves are constrained by $0 \leq \Delta\Pi \leq 1$, where $\Delta\Pi = 1$ represents a 100% power reduction. The position of a TCL class on this chart indicates the maximum power reduction that could be provided for secondary/flexible response, assuming initialization in the steady state and an ideal controller (no limitations on the instantaneous power excursions). Each class is represented by three symbols: the square specifies the coordinates of each representative appliance, highlighting its effective energy reduction capacity $(1 - \hat{\sigma}_{min})$ and its thermal relaxation time $\hat{\tau}$. The surrounding region covers the randomised appliances in its class and

the coloured dot represents the least flexible envelope model computed using equations (5.13)-(5.14). The ability for bottle coolers, upright fridges and freezers to provide long term services is significant as these appliances are placed above the upper envelope $\Delta\Pi = 1$. Hence, they can (in principle) remain switched off for more than 30 minutes. In contrast, the maximum contribution for the domestic refrigeration is limited to only a fraction of their aggregate power. Ultimately, the multidecks are not suitable for long term services due to their small thermal time constant $\hat{\tau}$, despite a considerable thermal headroom $(1 - \hat{\sigma}_{min})$.

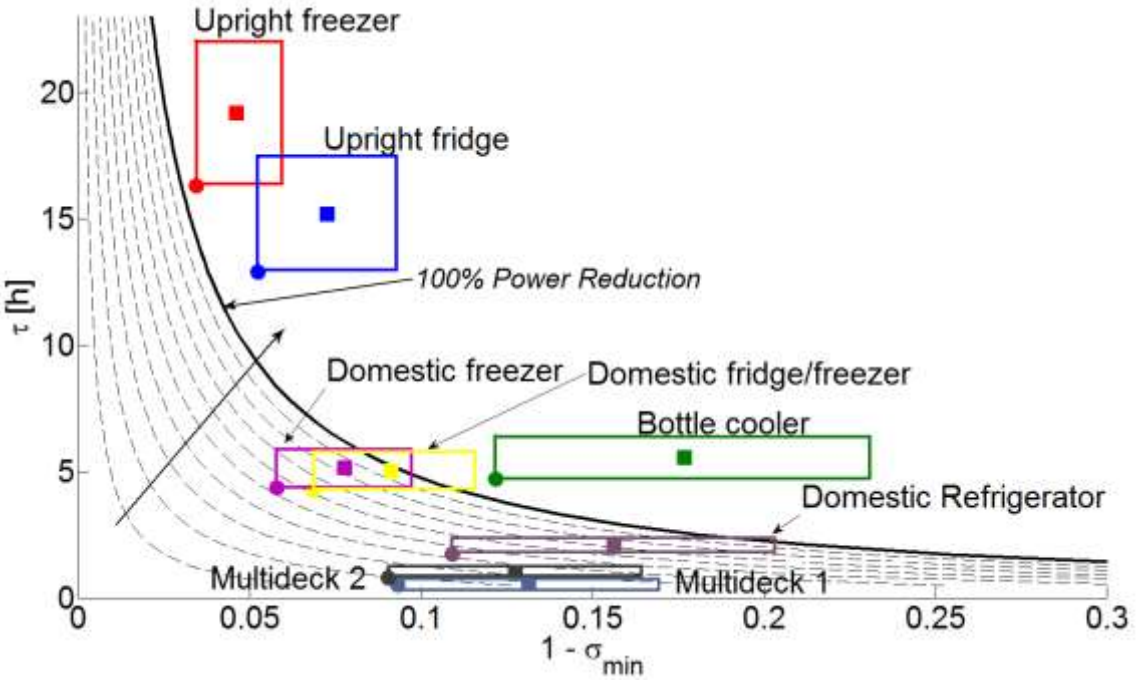


Figure 33 Ability to provide 30 minute demand reduction for secondary/flexible response. Each dot represents a class of appliances; dashed curves are the power reduction obtained by solving (5.16) for $\Delta\Pi = 1$. The arrow points in the direction of increased power reduction, saturating at 100%.

5.3.3 Energy arbitrage

The ability of TCLs to exploit their intrinsic energy reservoir to profit from energy arbitrage is now analysed. In order to estimate the TCLs' performance we consider how much energy can be provided or absorbed without violating the devices' energy limits; in other words, how much energy can effectively be 'transferred' between time windows. For this generic analysis the power constraints imposed by the controller are ignored and thus it is assumed that $\Pi(t) = \pi(t)/\pi_0 \in [0,1/\pi_0]$.

First equation (5.14a) is solved using $\sigma(0) = \hat{\sigma}_{max}$ and $\Pi(t) = 0$ (no power consumption), and we calculate the maximum time t^* that this reduction can be sustained so that $\sigma(t^*) = \hat{\sigma}_{min}$:

$$t^* = \hat{\tau}[\log(\hat{\sigma}_{max}) - \log(\hat{\sigma}_{min})] \tag{5.17}$$

The amount of effective energy ΔE provided is defined as the time t^* it takes to deplete the storage capacity multiplied by the reduction in power consumption relative to the steady state:

$$\Delta E = t^* \cdot \hat{P}_0 = \hat{\tau}\hat{P}_0\Delta\sigma + O(\Delta\sigma^2) \tag{5.18}$$

where $\Delta\sigma = \widehat{\sigma}_{max} - \widehat{\sigma}_{min}$ represents the dimensionless accessible storage capacity.

For the opposite case of energy absorption we consider the initial condition $\sigma(0) = \widehat{\sigma}_{min}$ and $\Pi(t) = 1/\pi_0$ (the maximum attainable value where all devices are forced on), resulting in

$$t^* = \hat{\tau}\left[\log\left(1 + \frac{\pi_0(1-\widehat{\sigma}_{min})}{1-\pi_0}\right) - \log\left(1 + \frac{\pi_0(1-\widehat{\sigma}_{max})}{1-\pi_0}\right)\right] \tag{5.19}$$

The energy stored is the time at full capacity t^* times the increase in power consumption relative to the steady state $\left(\frac{\hat{P}_0(1-\pi_0)}{\pi_0}\right)$. An expansion in powers of $\Delta\sigma$ again results in

$$\Delta E = \hat{\tau}\hat{P}_0\Delta\sigma + O(\Delta\sigma^2) \tag{5.20}$$

The approximate energy transfer capacity per installed capacity Δe (in units of P_0) is therefore

$$\Delta e = {}^{\Delta E}/_{\hat{P}_0} \cong \hat{\tau} \cdot \Delta\sigma \quad [MWh/MW] \tag{5.21}$$

Figure 34 shows the appliance types on the $(\Delta\sigma, \hat{\tau})$-plane. The dashed curves represent iso-lines of the approximate energy arbitrage capacities divided by $\hat{P}_0$. As before, the squares (representative appliances), boxes (heterogeneous populations) and dots (least flexibility envelope) indicate the abilities of particular appliance classes.

The highest potential for energy arbitrage is gathered in the upper-right side of Figure 34, where a large usable storage capacity is used in conjunction with a large thermal relaxation time. Mirroring the medium term services, the commercial upright fridges and freezers and the bottle coolers are the most suitable units for energy arbitrage.

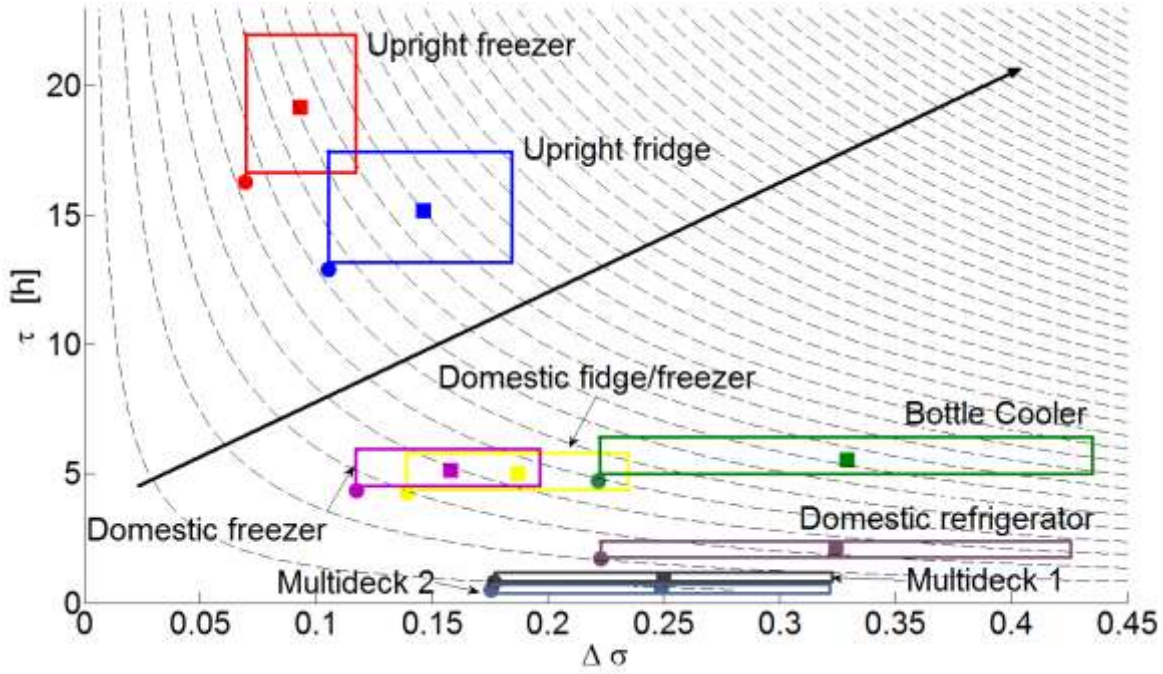


Figure 34 Approximate energy arbitrage capacity. Each dot represents a class of appliances. Dashed lines represent usable storage capacities; the arrow points in the direction of higher capacities.

As explained so far, the position of the squares representing different classes of TCLs suggests the high or low ability of each individual class of devices to provide a particular system service. For this purpose, Figures 33-34 suggest that different classes of appliances could be clustered and create a unique larger class of TCLs; in particular domestic refrigerators and the two multideck units exhibit similar performances regarding the provision of secondary/flexible response (Figure 33) and energy arbitrage (Figure 34). Same considerations are extended to domestic fridge/freezers and domestic freezers. Clustering classes of devices that are in principle different would help TSOs or demand side aggregators in order to reduce the dimension of the problem. Moreover, it is worth pointing out that domestic refrigerators and multideck share similar performances only with regard to long term services whereas, as shown in Figure 32, the ability to provide short-term primary or high frequency response differs significantly.

5.4 The Leaky storage unit

The representation using dimensionless parameters Π and σ is particularly attractive for analysing the behaviour of heterogeneous devices. However, from the power system perspective it is more instructive to reinterpret the envelope model in equations (5.13) and (5.14) as an effective energy storage unit. Defining the energy level $S(t)$ as

$$S(t) \equiv \hat{P}_0 \hat{\tau} \sigma(t) \tag{5.22}$$

and using equation (5.13c), equation (5.14a) may be written as

$$\frac{dS(t)}{dt} = -\frac{1}{\hat{\tau}} S(t) + P(t). \tag{5.23}$$

This offers an interpretation of the aggregate TCLs as a *leaky storage unit*. This storage unit loses its stored energy with an intrinsic rate $\hat{\tau}^{-1}$ and it is replenished by the variable power consumption $P(t)$. It is worth pointing out that the leaky storage unit cannot physically discharge electricity; it is only able to decrease or increase its consumption compared to its steady state consumption $\hat{P}_0$.

The operational limits in Equations (5.13a) and (5.13b) can be transformed to explicit power and energy constraints of the form

$$P_{min} \leq P(t) \leq P_{max} \tag{5.24a}$$

$$S_{min} \leq S(t) \leq S_{max} \tag{5.24b}$$

using (5.13c) and (5.22). The range between the energy bounds may be interpreted as a usable storage capacity:

$$\Delta S = S_{max} - S_{min} \tag{5.25}$$

Finally, the steady state energy level is given by $S_0 = \hat{P}_0 \hat{\tau}$.

An intuitive image of this leaky storage model is that of a column of water, depicted in Figure 35. A faucet controls the influx of water ($P(t) \geq 0$) and water leaks out through a hole at the bottom. The flow through the faucet must be controlled in such a way that the water level in the column remains between specified minimum and maximum levels.

The provision of demand response services should not excessively impact the primary cooling/heating function of TCLs. Temperature bounds alone are insufficient for this purpose, as appliances could opt to spend all time at the lower or upper temperature bounds, clearly affecting the primary function of these devices. Therefore, the expected temperature fluctuations of each appliance is forced to average to zero across a time window of interest w.

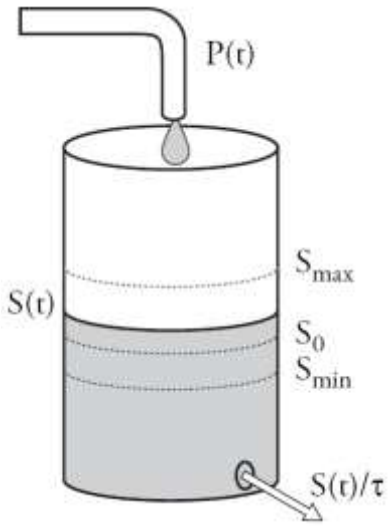


Figure 35 The leaky storage unit

Because population-average temperature and storage level are linearly related by equations (5.22) and (5.6) this implies

$$\frac{1}{w}\int_{w} S(t)\mathrm{dt} = \mathrm{S}_0 \tag{5.26}$$

By combining this with (5.23) it follows that the average power consumption must also equal the steady state value $\hat{P}_0$.

5.5 Optimal allocation of energy arbitrage and frequency services

The high level model of TCLs as a leaky storage unit enables a straightforward embedding in a linear optimisation framework for the simultaneous allocation of multiple demand response services. In the following, two distinct types of services are considered. The first is energy arbitrage, which enables the TCLs to benefit from energy price variability between different hours of the day. The second type refers to frequency services that support the power system balancing. For the latter we consider only frequency support for infrequent severe events, as opposed to real-time balancing services. The ability of the TCL controller to track an arbitrary reference profile $P(t)$ makes it possible to deliver both energy arbitrage and frequency services simultaneously, so long as their simultaneous provision does not exceed the physical constraints of the appliances (see equation (5.24)).

In the optimal service allocation problem the two services are treated in distinct ways. Committing TCLs to perform energy arbitrage implies actual changes in the energy profile. This is in contrast to the provision of frequency services, for which the TCLs receive an availability fee even when the service is not called upon. In the following we will assume that activation of frequency services is rare, so that the costs and fees associated with this utilisation are negligible. As a result, the only important consideration for frequency services is that the TCLs maintain sufficient power and energy margins at all times to ensure the deliverability of the frequency services they have committed to, while operating within set temperature limits.

In Section 5.4 the quality of service temperature bounds have been represented by the energy bounds S_{max} and S_{min}. In practice, it is likely that different bounds will be used for regular (i.e. energy arbitrage) and emergency (i.e. frequency response) services, reflecting the need to maintain stricter temperature limits in everyday use. The full permissible temperature range would then only be used for rare severe frequency events. In the optimisation framework this will be reflected by differentiated energy bounds as follows:

$$S_{min} < S_{min}^{r} \leq S(t) \leq S_{max}^{r} < S_{max} \tag{5.27a}$$

$$S_{max} = S_0 + r_{\%} \cdot (S_{max} - S_0); \quad S_{min} = S_0 - r_{\%} \cdot (S_0 - S_{min}); \tag{5.27b}$$

where $r_{\%}$ is the percent amplitude of the of the energy range. The frequency services considered provide system support after a severe power imbalance. The included services are modelled according to the National Grid practice: primary response, high frequency response, secondary and flexible response [78] (described in Section 5.3). The allocated response for the previous services is constant for the entire period under consideration (in our case 24h). The flexible response is equivalent to secondary response but may be called upon only within two predefined temporal windows $w_1$and w_2 (typically morning and evening). Note that the amount of response provided within each window is also constant.

5.5.1 The optimization problem

Let us now consider the optimal allocation of multiple services on a generic day, represented by a 24 hour period (w=24 hours) with periodic boundary conditions. The optimisation horizon divided into $m = 48$ periods i of duration Δt = 30 minutes. Solving (5.23) on the interval $[0, t]$ and

imposing constant electrical power consumption P_i within each interval *i* results in the discrete energy evolution equation:

$$S_{i+1} = S_i \cdot e^{-\frac{\Delta t}{\hat{\tau}}} + \hat{\tau} P_i \cdot \left(1 - e^{-\frac{\Delta t}{\hat{\tau}}}\right) \tag{5.28}$$

where S_i and S_{i+1} are the energy levels [MWh] at the beginning and end of interval i, respectively. We choose to formulate the problem so that the set of discrete energy levels $\mathcal{S} = \{S_i\}_{i=1}^{m}$ are decision variables. The corresponding discrete power levels P_i can be computed by inverting (5.28), made explicit by the notation $P_i(\mathcal{S})$. TCLs are modelled collectively as a price-taking leaky storage unit, so the optimisation takes the form of payment minimization problem that considers the cost of electricity consumption minus any fees for frequency services. It is assumed that prices are known at the time the optimisation is performed. The decision variables are the energy levels $\mathcal{S}$ [MWh] and the committed frequency services P^p (primary response), P^h (high frequency response), P^s (secondary response) and P^{f1} & P^{f2} (flexible response) [MW].

$$\min_{\mathcal{S},\mathrm{P^p},\mathrm{P^s},\mathrm{P^h},\mathrm{P^f}} \left\{ \sum_{\mathrm{i=1}}^{\mathrm{m}} \rho_\mathrm{i} \cdot \mathrm{P_i}(\mathcal{S}) \cdot \Delta\mathrm{t} - \mathrm{h^f} \cdot \left(\mathrm{w_1} \cdot \mathrm{P^{f1}} + \mathrm{w_2} \cdot \mathrm{P^{f2}}\right) - \mathrm{w}\left(\mathrm{h^p} \cdot \mathrm{P^p} - \mathrm{h^h} \cdot \mathrm{P^{hf}} - \mathrm{h^s} \cdot \mathrm{P^s}\right) \right\} \tag{5.29}$$

Subject to (for all i):

$$P_{min} \leq P_i(\mathcal{S}) \leq P_{max} \tag{5.30}$$

$$S_{min}^r \leq S_i \leq S_{max}^r \tag{5.31}$$

$$\frac{1}{m}\sum_{i=1}^{m} S_i = S_0 \tag{5.32}$$

$$0 \leq P^p \leq P_i(\mathcal{S}) - P_{min} \tag{5.33a}$$

$$0 \leq P^h \leq P_{max} - P_i(\mathcal{S}) \tag{5.33b}$$

$$P^s \geq 0 \tag{5.34a}$$

$$P^{f1} \geq 0 \tag{5.34b}$$

$$P^{f2} \geq 0 \tag{5.34c}$$

$$P^s + P^{f1} \cdot \mathbb{I}_{i \epsilon w_1} + P^{f2} \cdot \mathbb{I}_{i \epsilon w_2} \leq P_i(\mathcal{S}) - P_{min} \tag{5.35a}$$

$$P^s + P^{f1} \cdot \mathbb{I}_{i\epsilon w_1} + P^{f2} \cdot \mathbb{I}_{i\epsilon w_2} \leq P_{i+1}(\mathcal{S}) - P_{min} \quad (5.35b)$$

$$S_{i+1} - \hat{\tau}\left(1 - e^{-\frac{\Delta t}{\hat{\tau}}}\right) \cdot \left(P^s + P^{f1} \cdot \mathbb{I}_{i\epsilon w_1} + P^{f2} \cdot \mathbb{I}_{i\epsilon w_2}\right) \geq S_{min} \quad (5.36a)$$

$$S_{i+2} - \hat{\tau}\left(1 - e^{-\frac{\Delta t}{\hat{\tau}}}\right) \cdot \left(P^s + P^{f1} \cdot \mathbb{I}_{i\epsilon w_1} + P^{f2} \cdot \mathbb{I}_{i\epsilon w_2}\right) \geq S_{min} \quad (5.36b)$$

Note that the electricity price ρ_i per period i and the availability fees h^x for each frequency service x are expressed in £/MWh. The solution is bounded by the controller's power constraints (5.30) and the energy constraints for regular operation (5.31). In addition, (5.32) enforces the average energy (temperature) to equal the steady state energy (temperature). This constraint follows from inserting (5.23) in (5.26) and using (5.28) to eliminate P_i in each interval. We refer the reader to the appendix to this chapter for the mathematical derivation of (5.32). Adequate reserves for primary response and high frequency response allocations are enforced by (5.33). The same aim characterises (5.34-5.35), with regard to secondary and flexible response inside and outside w_1 (P^{f1}) and w_2 (P^{f2}). The indicator function $\mathbb{I}_s$ returns 1 if statement s is true, and 0 otherwise. It is assumed that the provision of primary and high frequency response for 30 seconds has a negligible impact on energy levels. A frequency event may occur at any time within an interval, and the secondary/flexible response commitments must be provided for 30 minutes from that moment onward. For this reason we must ensure that the services agreed for interval i can also be provided for the interval $i + 1$, resulting in constraint (5.35_b). Equations (5.36) ensure that calling upon the secondary and flexible frequency services will not conflict with the emergency lower energy limit S^e_{min}. Similar to the power constraint, the energy constraint is also enforced for the subsequent interval $i + 1$.

5.6 Case study and results

The considered half-hourly price profile is shown in Figure 36; it reflects the load and generator dispatch on the GB system on a representative winter day [79].

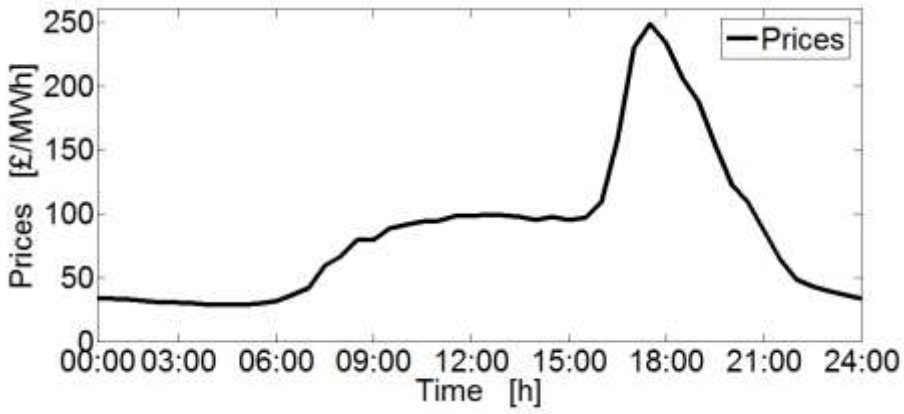


Figure 36 Price profile for a typical winter day

The response services are priced at h^p=6 £/MWh, h^h=7 £/MWh, h^s=5 £/MWh and h^f=8 £/MWh, in line with National Grid values [80]. The slightly higher value for h^f reflects the desire to incentivize the provision of response services within the two flexible response windows: w_1 ranges from hour 06:30 to 09:30 and w_1 from 16:00 to 20:00. For a given price profile equations. (5.29-5.36) fully define a linear optimisation problem that determines the optimal allocation of services that maximises the TCLs' profit across the 24 hour window.

Figure 37 shows the optimisation result for a heterogeneous population of domestic refrigerators (here 1MW base load is used for illustration); similar profiles characterise the other classes of TCLs. The energy profile of the leaky storage unit corresponding to the domestic refrigerators is shown in Figure 37A. The regularly accessible energy range is assumed to be $r_{\%}$ =75% with respect to S_0, of the full accessible range, in accordance with (5.27). The optimised energy level (solid black) is generally high in the morning in response to low prices during the night and mostly high during the afternoon as a result of high prices during the day. This profile is constrained by the regular energy limits and the average energy constraint (5.32). Figure 37B illustrates the corresponding optimal power policy (solid black) compared to the nominal power consumption (blue).

Figure 37C illustrates the allocation of frequency services. Supplying short term services, primary (green) and high frequency response (grey) restricts the power levels accessible for energy arbitrage, but is nevertheless desirable due to the high holding prices for these services.

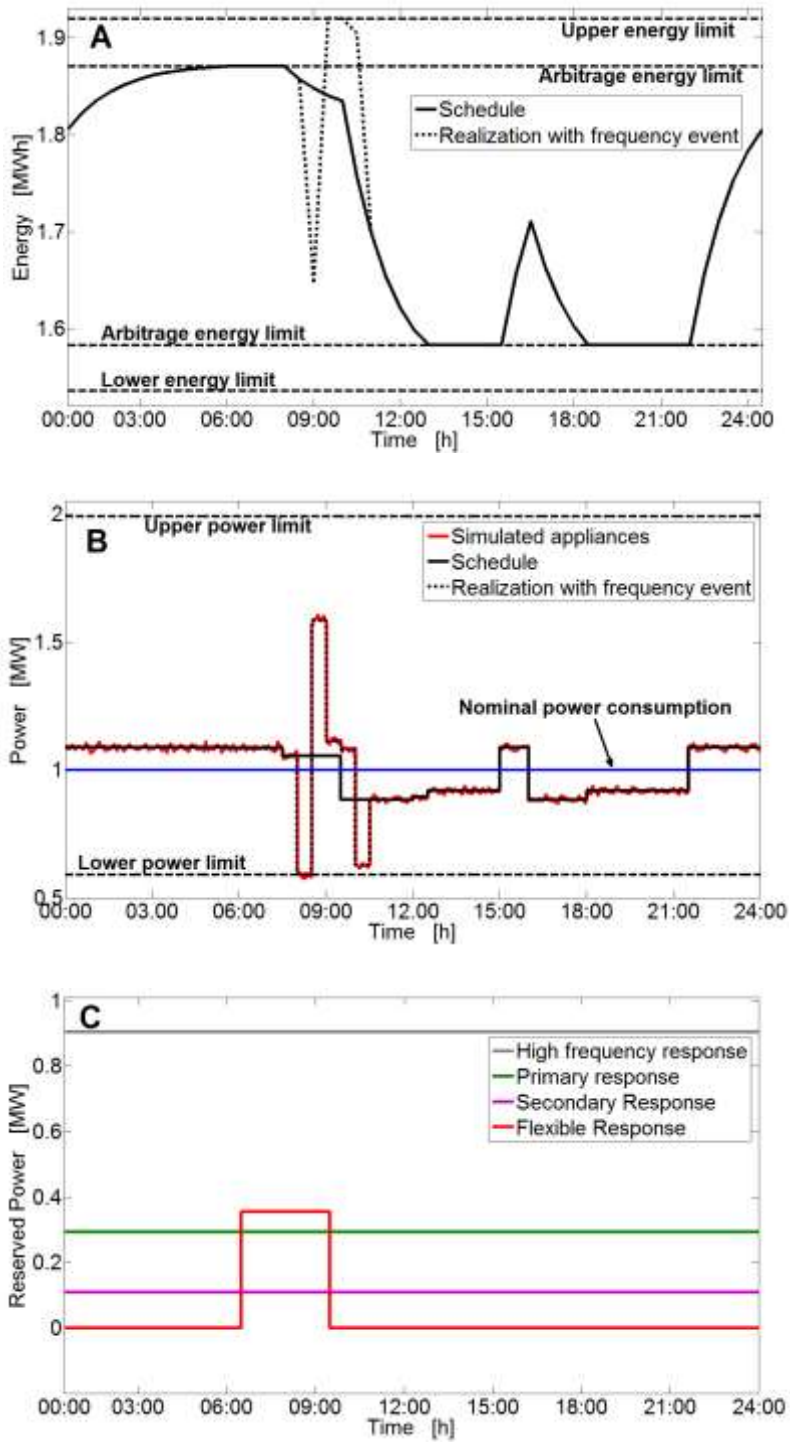


Figure 37 Optimal service allocation for 1MW of domestic refrigerators. (A) Optimal energy profile (solid) and perturbed energy profile (dotted) in case of provision of frequency services. (B) Optimal power profile (solid black) compared to the nominal consumption (blue). The dotted curve is the realised power consumption after provision of frequency services; Device-level simulation results based on this solution are shown in red (solid). (C) Optimal allocation of the frequency response services.

In contrast, the allocation of 30-minute secondary and flexible response services impacts both power and energy constraints, and as a result the committed amounts are lower. Moreover, further studies reveal that the secondary response allocation would be absent altogether if there

was no distinction between the regular and emergency energy limits. Lastly, flexible response availability (red) is selected only for w_1 facilitated by low electricity prices in this period. During w_2 the energy price reaches its highest value, so it is desirable to deplete the energy reserves and reduce the aggregate power consumption to a minimum. This conflicts with the provision of flexible response in this window.

The black dotted curve in Fig. 37A-B illustrates the deliverability of contracted ancillary services and the recovery of normal operations after their provision. In this example a sudden loss of generation activates the primary, secondary and flexible response services at 08:30. As a result the energy level drops and deviates from the scheduled profile. An optimal recovery pattern is computed to force the TCLs to return to the scheduled energy profile as fast as possible, subject to the average energy constraint (5.32). The return trajectory must respect the power constraint (5.30) and the energy constraints as in (5.31), but has access to the full energy range instead of the restricted range; this relaxation facilitates the rapid return to the predetermined energy profile. The underlying assumption is that the system holds sufficient reserve generators available to supply the demand peak[5]. Note that the average energy constraint (5.32) is quite unlike a regular storage unit, as it enforces a period of 'overcharging' after the initial recovery. Finally, it is worth pointing out that different recovery policies can be implemented depending on TSO requirements and commercial arrangements.

To demonstrate the practical feasibility of the resulting power profile, Fig. 37b also includes the simulated power consumption of 59,524 (1MW/$\hat{P}_0$) domestic refrigerators with randomised parameters according to Table 9. The appliances were controlled in a fully decentralised manner according to the control strategy and algorithm Chapter 4, with modifications (see equation (5.21)) to incorporate the regime switching mentioned in Section 5.2.2.

Differences between TCLs' physical models result in different optimal allocation of energy arbitrage and frequency services. The profit calculation in Figure 38 highlights the particular dynamics of each class of device; each colour represents a different component of the daily profit of TCLs. The profit derived from the provision of a frequency service is the product of the power reserved for that service (Figure 37C), the corresponding availability fee and the applicable service

[5] This assumption would be removed in Chapter 7 where the impact of TCLs' recovery on the reserve generators' will be considered in detail.

window. The profit due to energy arbitrage is calculated as the additional expense incurred by TCLs if they would consume the nominal power.

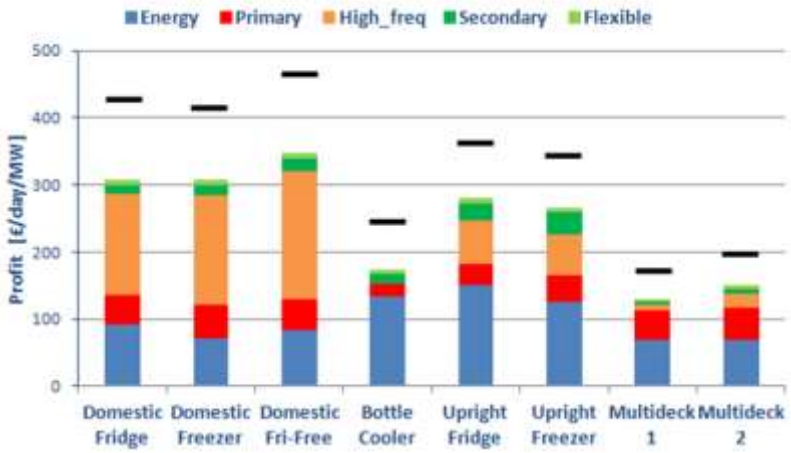


Figure 38 Daily profit per class of TCLs realised by energy arbitrage and by selling availability for frequency services. Solid markers indicate the total profit for homogeneous TCL populations.

The results largely mirror the ordering shown in Figures 32-34, adjusted by the price-driven competition between services. Domestic appliances make most of their profit from the provision of high frequency response (orange) followed by energy arbitrage (blue), whereas other services contribute only marginally. In contrast, upright fridges and freezers can profit more from energy arbitrage than from high energy services, reflecting their lower power headroom (Figure 32) and larger thermal time constants (Figure 34). The bottle coolers and multidecks are limited to smaller potential profits, resulting primarily from their very limited power headroom (Figure 32) and – for multidecks – their weak thermal insulation (Figures 33-34). Finally, the markers in Figure 38 indicate the total profits of hypothetical homogeneous populations where all TCLs are described by the representative device parameters in Table 9. The gap between the markers and the top of the bars is indicative of the loss of profit due to the aggregation of heterogeneous appliances in a single control cluster.

5.7 Conclusions

In this chapter we have demonstrated how a large heterogeneous collection of thermostatically controlled loads can be modelled as a leaky storage unit, a high level representation that is particularly insightful from the power system perspective. The accessible storage levels $S(t)$ and power consumption levels $P(t)$ of this unit are limited by the appliances'

thermal models, their temperature constraints and the control strategy adopted. The model has been introduced in the context of cold appliances, but it applies equally to heating appliances. The leaky storage model is exact for large populations of TCLs ($N \to \infty$) with linear thermal models that are controlled by the control framework discussed in the last chapter. In addition, following the extensions to the basic version of the two-stage controller (see Section 4.9.1), we matched the basic controller, designed for power-reduction only, with its power-increasing counterpart. In this way we achieved full use of the available temperature interval.

Moreover, individual appliances are not identical. Hence, the envelope equations (5.13-5.14) also guarantee the deliverability of a desired response by a heterogeneous collection of appliances. The envelope equations may also be used to increase the robustness of solutions to unknown variations in model parameters. Even when the true appliance parameters are not accurately known when services are allocated, the appliances will still be able to provide the desired response provided that

1) Each appliance has learned its true model by the time the response is executed.

2) The envelope encompasses the realised models of all appliances.

This will be the case even when the models change over time (e.g. day-night cycle), as long as the change in model parameters is sufficiently slow compared to the appliances' intrinsic dynamics. The leaky storage representation allows for an embedding in familiar revenue-maximising scheduling models. This enables an optimal allocation of energy and ancillary services that takes into account (a) the physical properties, (b) the constraints associated with the quality of the service of the appliances, and (c) any conflicts or synergies between different services. In this chapter we have demonstrated this potential with the optimal allocation of energy consumption and various frequency services for a typical GB winter day. The services modelled are primary and high frequency response (short term services) and secondary and flexible response (long term services) The results obtained in Section 5.6 for eight classes of refrigeration units show that very different service allocations result from the differences in physical models. The allocations can largely be understood from the analysis carried out by means of dimensionless parameters introduced in Section 5.3. Moreover, the behaviour of the appliances also reflect the price-driven competition between services. Furthermore, the results show the benefit to aggregators of clustering similar appliances, and the reductions in profit that may result from diverse TCL models within a cluster (Figure 38).

5.7.1 Extensions and links with the next chapters

The results presented in this chapter suggest a number of pathways for further development. Initial extensions of the control strategy have been already considered but further steps may enable the use of more complex second order thermal models. Furthermore, the optimal service allocation case study may be enhanced in a number of ways. It currently assumes that responsive TCLs are not sufficiently numerous to affect energy and ancillary service prices. However, the future grid is expected to feature significant levels of TCL demand response, raising the possibility to exploit the collective flexibility offered by TCLs to relax constraints on the generators' dispatch. The exploration of this aspect is the subject of the next chapter; a security constrained economic dispatch model is developed to show the ability of TCLs demand response to reduce the operational cost of the system reducing the need for expensive peaking generators. The impact of the flexible TCLs consumption re-allocation is studied under the light of the ability to relieve transmission lines' congestions.

Moreover, the 24-hour periodic optimisation of service allocation could also be extended to an open loop optimisation on a longer time scale. This could take account the fact that the provision of ancillary services usually takes the form of longer-term contracts. Finally, the assumption that income and losses due to the utilisation of these services can be neglected may need to be reconsidered for future power systems as the frequency of utilisation may increase.

5.8 Appendix

APP5.1 Derivation of equality constraint (5.32)

The aim of this section is to validate constraint (5.32). Recalling equation (5.23), the general solution to this first order ODE is given by

$$S(t) = S_{in} \cdot e^{-\frac{t}{\tau}} + \int_0^t P(\theta) \cdot e^{-\frac{\theta - t}{\tau}} d\theta \qquad \text{(A5.1)}$$

Where S_{in} is the initial condition for the energy. Note that we do not use the 'hat' parameters as it is a general demonstration valid for both the cases, homogeneous or heterogeneous TCLs; Let us now calculate the integral of the energy $S(t)$ over a generic interval of length Δt.

$$\int_0^{\Delta t} S(t)dt \tag{A5.2}$$

Therefore, considering a constant power consumption $P(\theta) = P$ within this interval, we obtain:

$$\begin{aligned}
\int_0^{\Delta t} S(t)dt \; &= S_{in}\int_0^{\Delta t} e^{-\frac{t}{\tau}} + P\int_0^{\Delta t}\int_0^{t} e^{-\frac{\theta - t}{\tau}} d\theta \, dt = \\
&= -S_{in}\tau\left[e^{-\frac{t}{\tau}}\right]_0^{\Delta t} + P\tau\int_0^{\Delta t} e^{-\frac{t}{\tau}}\left[e^{-\frac{\theta}{\tau}}\right]_0^{t} dt = \\
&= -S_{in}\tau\left[e^{-\frac{\Delta t}{\tau}} - 1\right] + P\tau\int_0^{\Delta t}\left[1 - e^{-\frac{t}{\tau}}\right] dt = \\
&= S_{in}\tau\left(1 - e^{-\frac{\Delta t}{\tau}}\right) + P\tau\Delta t + P\tau^2\int_0^{\Delta t} -\frac{1}{\tau}e^{-\frac{t}{\tau}}dt = \\
&= S_{in}\tau\left(1 - e^{-\frac{\Delta t}{\tau}}\right) + P\tau\Delta t + P\tau^2\left(e^{-\frac{\Delta t}{\tau}} - 1\right) =
\end{aligned} \tag{A5.3}$$

We now express the power consumption P as function of the energy levels at the beginning (S_{in}) and the end (S_{fin}) of the interval considered. For this purpose, we rearrange equation (5.28).

$$P = \frac{S_{fin} - S_{in}e^{-\frac{\Delta t}{\tau}}}{\tau\left(1 - e^{-\frac{\Delta t}{\tau}}\right)} \tag{A5.4}$$

Hence considering (A5.3) and (A5.4) it results that,

$$\begin{aligned}
\int_0^{\Delta t} S(t)dt &= S_{in}\tau\left(1 - e^{-\frac{\Delta t}{\tau}}\right) + S_{fin}\frac{\Delta t}{\left(1 - e^{-\frac{\Delta t}{\tau}}\right)} - S_{in}\frac{\Delta t e^{-\frac{\Delta t}{\tau}}}{\left(1 - e^{-\frac{\Delta t}{\tau}}\right)} + S_{fin}\tau - S_0\tau e^{-\frac{\Delta t}{\tau}} = \\
&= S_{in}\left[\tau - \frac{\Delta t e^{-\frac{\Delta t}{\tau}}}{\left(1 - e^{-\frac{\Delta t}{\tau}}\right)}\right] + S_{fin}\left[\frac{\Delta t}{\left(1 - e^{-\frac{\Delta t}{\tau}}\right)} - \tau\right]
\end{aligned} \tag{A5.5}$$

The results obtained in (A5.5) is then divided by Δt so,

$$\frac{1}{\Delta t}\int_0^{\Delta t} S(t)dt = S_{in}\left[\frac{\tau}{\Delta t} - \frac{e^{-\frac{\Delta t}{\tau}}}{\left(1-e^{-\frac{\Delta t}{\tau}}\right)}\right] + S_{fin}\left[\frac{1}{\left(1-e^{-\frac{\Delta t}{\tau}}\right)} - \frac{\tau}{\Delta t}\right] \tag{A5.4}$$

The left hand side of (A5.4) is the definition of the energy average value over a generic interval of length Δt. For simplicity let us define

$$\varphi = \left[\frac{\tau}{\Delta t} - \frac{e^{-\frac{\Delta t}{\tau}}}{\left(1-e^{-\frac{\Delta t}{\tau}}\right)}\right] \tag{A5.5a}$$

$$\chi = \left[\frac{1}{\left(1-e^{-\frac{\Delta t}{\tau}}\right)} - \frac{\tau}{\Delta t}\right] \tag{A5.5b}$$

The aim of (5.32) is to guarantee that the average energy over the whole optimization horizon (24h) is equal to the steady state value S_0. Therefore, considering $m = 48$ periods i, all of them with duration $\Delta t = 30$ minutes (as in Section 5.5.1) and defining S_i and S_{i+1} as the energies at the beginning and the end of interval i respectively, the average energy over one day results in:

$$\frac{1}{m}\sum_{i=1}^{m}(\varphi S_i + \chi S_{i+1}) = S_0 \tag{A5.6}$$

Moreover the reader should note that:

$$\varphi + \chi = \frac{\tau}{\Delta t} - \frac{e^{-\frac{\Delta t}{\tau}}}{\left(1-e^{-\frac{\Delta t}{\tau}}\right)} + \frac{1}{\left(1-e^{-\frac{\Delta t}{\tau}}\right)} - \frac{\tau}{\Delta t} = 1 \tag{A5.7}$$

Therefore, it is easy to check that the equality constraint (A5.6) becomes

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}(\varphi S_i + \chi S_{i+1}) &= \varphi S_1 + (\varphi+\chi)S_2 + \cdots + (\varphi+\chi)S_m + S_{m+1} \\ &= \frac{1}{m}\left(\varphi S_1 + \sum_{i=2}^{m} S_i + \chi S_{m+1}\right) = S_0\end{aligned} \tag{A5.8}$$

Finally, the application of periodic boundary conditions in Section 5.5.1 implies that

$$S_1 = S_{m+1} \implies \frac{1}{m}\sum_{i=1}^{m}(\varphi S_i + \chi S_{i+1}) = \frac{1}{m}\sum_{i=1}^{m} S_i = S_0 \tag{A5.9}$$

Equation (A5.9) confirms the correctness of the equality constraint (5.32).

■

APP5.2 Derivation of equality constraint (5.32)

This section demonstrates a result announced in Chapter 3 and touched also in Chapter 4. The introduction of the leaky storage model in this Chapter facilitates the understanding this conclusion. Considering Figure 13, we found out that not all the borrowed energy needs to be paid back; this implies that the area A can actually be larger than the area B. The energy payback terminates when the average temperature $\bar{T}(t)$ of the TCL population achieves and maintains the steady state value T_0. However, this condition can be achieved in two ways:

$$\bar{T}(t) = T_0 \text{ and } \frac{1}{w}\int_w \bar{T}(t)dt < T_0 \tag{A5.10a}$$

$$\bar{T}(t) = T_0 \text{ and } \frac{1}{w}\int_w \bar{T}(t)dt = T_0 \tag{A5.10b}$$

We now want to demonstrate that the energy to payback (area B) is smaller than the energy borrowed (area A) if (A5.10a) is applied or rather if the mean of the average temperature over the time window of interest w is smaller than the steady state value T_0. Rearranging (5.22) considering (5.6b) we obtain the link between the average thermal energy and the average temperature;

$$\bar{T}(t) = T_{off} - \frac{S(t)\cdot\left(T_{off} - T_{on}\right)}{P_{max}\tau} \tag{A5.11}$$

Hence, with appropriate substitutions, we obtain

$$\frac{1}{w}\int_w \bar{T}(t)dt < T_0 \;\rightarrow\; T_{off}\frac{w}{w} - \frac{\left(T_{off} - T_{on}\right)}{P_{max}\tau}\frac{1}{w}\int_w S(t)dt < T_{off} - \frac{S_0\cdot\left(T_{off} - T_{on}\right)}{P_{max}\tau} \tag{A5.12}$$

$$\frac{1}{w}\int_w \bar{T}(t)dt < T_0 \;\rightarrow\; \frac{1}{w}\int_w S(t)dt < S_0 \tag{A5.13}$$

The result introduced with (A5.13) demonstrates that thermal energy borrowed from the TCLs (area A) is equal to the energy that TCLs have to payback (area B) to achieve a steady state condition $\bar{T}(t) = T_0$ only if the average temperature is also respected ($\frac{1}{w}\int_w \bar{T}(t)dt = T_0$) over the time window w. In this case also the average power consumption $S(t)$ must equal the steady state value P_0. This can be easily demonstrated by combining (5.26) with (5.23).

■

Chapter 6: Security Constrained Economic Dispatch with Flexible Thermostatic Loads

Abstract

Thermostatically controlled loads such as refrigerators have been proved to be natural candidates for short term demand response. The analysis carried out in this chapter builds on the results achieved in the previous chapters regarding the accurate aggregate control of a heterogeneous cluster of thermostatic loads described as a leaky storage unit. Hence, we take a step forward and change the point of view of the analysis; our technical results are exploited in order to quantify the value for the power system associated with the thermostatic loads' ability to (a) perform energy arbitrage, (b) to provide frequency response and (c) transmission constraint management services. The devices' ability to provide the first two types of services has been introduced in the previous chapters, whereas economic aspects associated to the last service are discussed in this chapter. We incorporate a multi-services demand response model in a Security Constrained Economic Dispatch (SCED) that minimizes the system operational cost of a two bus-bar system, subject to frequency response and transmission constraints. Further sensitivity studies assess the impact of different penetration levels of controllable loads and transmission flow constraints.

Nomenclature

The main variables and parameters used in this chapter are given below; others are provided throughout the chapter.

Variable	Description	Unit
w	*Optimization horizon*	h
i	*Time period index*	
g	*(superscript) Generator class*	
a	*(superscript) TCLs appliance class*	
P_i^g	*Generator power at interval i*	MW
$P_{d,i}^a$	*Aggregate power of TCLs class at interval i*	MW
$P_{s,i}$	*Static load interval at interval i*	MW
π_i^g	*Generator primary response at interval i*	MW
π_i^a	*Primary response of TCLs class at interval i*	MW
σ_i^g	*Generator frequency response at interval i*	MW
σ_i^a	*Frequency response of TCLs class at interval i*	MW
$\mathcal{G}$	*Set of all generator classes*	
A	*Set of all TCL classes*	
$\mathcal{G}_N$	*Set of generators in the northern node*	
$\mathcal{G}_S$	*Set of generators in the southern node*	
A_N	*Set of TCLs in the northern node*	
A_S	*Set of TCLs in the southern node*	

6.1 Introduction

In the last chapter we illustrated how thermostatically controlled loads (TLCs), such as refrigerators, are able to shift their aggregate power consumption in response to network requirements. This natural flexibility is enabled by slightly modifying the regular *on*/*off* operating cycle of each unit as long as the average temperature is approximately maintained over time; this way the primary function of these devices is not compromised. The ability to discharge/recharge the thermal energy stored in a cluster of TCLs to allow aggregate power reduction/increase, could

be strategically used by system operators to enhance the operation and security of future power system characterized by a substantial lack of flexibility in the generation side.

The ability of responsive refrigerators to provide inertial support (Chapter 2) and frequency control support has been discussed previously discussed (Chapter 2 and Chapters 3 and 4, respectively). These applications regard system security; an overall reduction in power consumption in response to a sudden frequency drop is the result of a generator outage. In addition, the adjustment of the nominal duty cycle of TCLs can also be employed for energy arbitrage, absorbing more energy when the electricity prices are low and less when high. This aspect has been discussed in Chapter 5.

This chapter aims to understand how the flexibility of a cluster of TCLs impacts the generator dispatch[6]. In one case, TCLs would be able to reduce the power consumption when the generation mix at a particular hour is very expensive and shift the demand to adjacent periods. If the power decrease is sufficiently large, this will change the dispatch of the peaking generators and these high-cost plants could be possibly de-committed. Moreover, responsive loads could also alleviate generators' ramping constraints. The generators dispatch would be altered also by means of the frequency response services from TCLs; in this case these plants would work more efficiently.

The ability to accurately control the aggregate TCLs power consumption over different time scales allows to achieve the aim of this chapter. Previous studies in this field had to depend on centralised decision models that rely on complex and costly real-time communication infrastructures [47]. Moreover, the leaky storage model offers a novel representation of the TCLs that is very convenient for system-level studies carried out in this chapter.

In this chapter, we quantify the value that TCLs' intrinsic flexibility has for system operation. This objective is analysed by means of a security constrained economic dispatch (SCED) model that optimizes the generators' output and TCL energy consumption in a two bus-bar system, subject to frequency response requirements and network power flows requirements. Moreover the flexible TCLs consumption enables the provision of transmission constraint management service; if the tie line connecting the two nodes of the system is congested, TCLs could facilitate the alleviation of the congestion. In this context, we study the potentially different TCLs behaviour depending on the node that these devices are connected to. Overall, the optimal trade-off between the system services

[6] In this chapter the TCLs aggregate power consumption is no longer reduced to a small fraction 1 MW.

minimizes the system operation cost. Although system-level economic benefits of dynamic demand have been already quantified [46], the support from TCL was limited only to primary frequency response.

The rest of the paper is organized as follows: Section 6.2 briefly summarizes the key results on aggregate TCLs modelling and control previously discussed. Section 6.3 describes the SCED model, including constraints related to TCLs participation. Moreover, Section 6.4 presents the framework and the assumptions for the case study, while the results of the optimization problem are presented in Section 6.5; in this part we also perform further sensitivity studies regarding the load flow congestions and CO_2 emissions curtailment. Finally, section 6.5 concludes the chapter summarizing the analysis carried out and offering the links with the next chapter.

6.2 Aggregate TCLs as a Leaky Storage Unit and control strategy

In this section we briefly recall the results obtained in Section 5.4. We have demonstrated that a large population of $N \gg 1$ of TCLs[7] can be described as an energy storage unit where the energy level S(t) is defined as:

$$\frac{dS(t)}{dt} = -\frac{1}{\tau}S(t) + P(t) \tag{6.1}$$

with $P(t) \equiv P_0 \Pi(t)$ where P_0 is the aggregate steady state power consumption and $\Pi(t)$ is a relative power curve. Moreover, the correspondence between energy level and the average temperature $\bar{T}(t)$ of the population of appliances is specified as:

$$S(t) \equiv P_{max} \cdot \left(\frac{T_{off} - \bar{T}(t)}{T_{off} - T_{on}}\right) \cdot \tau \tag{6.2}$$

with P_{max} representing the maximum aggregate power consumption. In addition, temperature constraints on individual appliances impose energy bounds on the storage capacity as for equation (5.27); in particular $[S^r_{min}, S^r_{max}]$ for regular operation (arbitrage) and $[S_{min}, S_{max}]$ for response services. The energy storage level respects:

$$S_{min} < S^r_{min} \leq S(t) \leq S^r_{max} < S_{max} \tag{6.3}$$

[7] For simplicity, we now assume a homogeneous population of TCLs but extensions to heterogeneous clusters are straightforward considering Section 5.2.3.

In addition it is reasonable to ensure that the population average temperature remains close to its steady state value T_0 on average. This is expressed by the following integral constraint across a time window of duration w

$$\frac{1}{w}\int_{w} S(t)\mathrm{dt} = \mathrm{S}_0 \tag{6.4}$$

The decentralised control method introduced in Chapter 4 enables each TCL appliance to independently target a relative power curve $\Pi(t)$ so that the aggregate power consumption tracks such a profile in expectation:

$$E[P(t)] = P_o \Pi(t) \tag{6.5}$$

Besides the generic limitations on energy (temperature), the controller implementation will limit the accessible range of power consumption levels. This implicitly defines the minimum and maximum power limits

$$P_{min} \leq P(t) \leq P_{max} \tag{6.6}$$

with

$$P_{min} = P_0 \cdot \frac{(T_{off} - T_{min})(\bar{T}_o - T_{max})}{(T_{min} - T_{max})(T_{off} - \bar{T}_o)} \tag{6.7}$$

$$P_{max} = P_0 \frac{(T_{max} - \bar{T}_o)(T_{max} - T_{on}) + (T_{off} - T_{max})(T_{max} - T_{min})}{(T_{min} - T_{max})(T_{off} - \bar{T}_o)} \tag{6.8}$$

This control method guarantees the feasibility of the response if (6.3) and (6.6) are satisfied, avoiding the need for device-level simulations. Hence, provided that a particular device can track a reference power profile $P(t)$ in expectation, the whole population can do so by extension. Under this strategy, TCLs can manage the actual energy consumption (energy arbitrage) and deliver frequency services simultaneously using flexible power curves $P(t)$, so long as their simultaneous provision does not violate device-level constraints.

6.3 Security constrained economic dispatch: model and assumptions

We study the impact of flexible TCL demand on a simple two bus-bar system model that mimic the GB scenario; however the simplicity of the model helps to better understand the response of flexible loads.

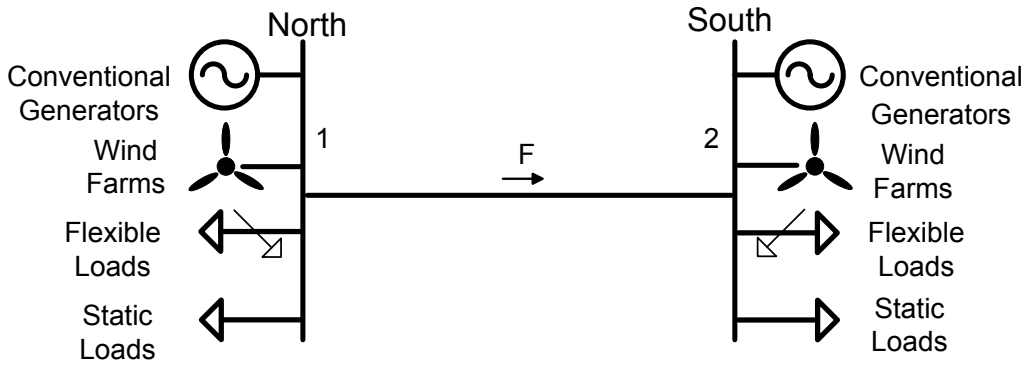


Figure 39 Schematic power system model

The model includes eight different generation technologies. The northern node (Scotland) is connected to the southern node (England and Wales) through a constrained tie line. The direction of the flow direction is as in Figure 39 due to a low demand and significant wind share in the northern node. Moreover, loads are divided in flexible loads (TCLs) and inflexible loads (generic demand assumed to be constant at each time interval). The TCLs may be used not only to make energy arbitrage but also to support the fulfilment of frequency services requirements. The modelled services are primary and secondary response as defined in the GB context [78]. The primary response service requires a decrease/increase in power consumption within 10s that is sustained for 20s, while secondary response imposes a decrease/increase in power consumption for up to 30 minutes that is provided within 30s. The commitment of these services is now assumed flexible, therefore different amounts of response services can be provided at different hours. For simplicity we do not consider the high frequency and flexible response services [78].

This SCED model secures the system response up to 30 minutes after a generation outage; therefore we consider this time span sufficiently ample to allow other generators to be committed to provide the required amount of reserve power and replace those machines and loads that were supplying secondary control. In the next chapter, a more detailed unit commitment model will take into account the increased provision of reserve services due to the TCLs energy/power recovery. Furthermore, we assume that the installed capacity of each generation technology is much larger

than the size of the individual machines within each group; therefore a coefficient $\varphi \in [0,1]$ determines the fraction of machines of each technology that is committed, rather than using integer decisional variables associated to individual units. Finally, we assume that wind output is perfectly forecast.

6.3.1 The SCED optimization model

We now proceed to formulate the optimization model for the security constrained economic dispatch; the aim of the problem is the minimization of the system operation cost exploiting the flexibility provided by responsive TCLs. We consider a time horizon w=24 hours, divided into m=48 periods i of Δt = 30 minutes with periodic boundary conditions. Solving (6.1) and imposing constant electrical power consumption within each interval, the discrete energy evolution results in the following equation:

$$S_{i+1} = S_i \cdot e^{-\frac{\Delta t}{\tau}} + \tau P_i \cdot \left(1 - e^{-\frac{\Delta t}{\tau}}\right) \tag{6.9}$$

It is worth pointing out that S_i is the energy level [MWh] at the beginning of the correspondent interval and S_{i+1} equals the energy level at the end of previous interval, i. Once defined the set of energy levels $\mathcal{S} = \{S_i\}_{i=1}^{m}$ this expression calculates the discrete power profile $\{P_i\}$. We adopt a quadratic cost function for the generators' power production; it is composed of a fixed term α_g [£/h], a linear coefficient β_g [£/MWh] and a quadratic term γ_g [£/(MWh)2]. The use of a quadratic cost function is consistent with several works in literature that deal with the economic dispatch problem (e.g. [81, 82]). This choice does not conflict with the embedding of the leaky storage model and associated constraints as the demand side model exhibits a straightforward implementation in any optimization tool. The SCED problem is therefore formulated as follows:

$$\min_{F,P,\mathcal{S},\pi,\sigma} \sum_{i=1}^{m} \sum_{g=1}^{\mathcal{G}} \left(\Delta t * \alpha_g * \varphi_i^g + \Delta t * \beta_g * P_i^g + \Delta t^2 * \gamma_g * P_i^{g^2}\right) \tag{6.10}$$

Subject to (for all i):

System level constraints:

$$-F_{max} \leq F_i \leq F_{max} \tag{6.11}$$

$$\sum_{a\in A_N} P_d(\mathcal{S})_i^a - \sum_{g\in \mathcal{G}_N} P_i^g + F_i + P_{s,i}^{(N)} = 0 \quad (6.12a)$$

$$\sum_{a\in A_S} P_d(\mathcal{S})_i^a - \sum_{g\in \mathcal{G}_S} P_i^g - F_i + P_{s,i}^{(S)} = 0 \quad (6.12b)$$

$$\sum_{g\in \mathcal{G}} \pi_i^g + \sum_{a\in A} \pi_i^a \geq R_{max} + K_i \cdot \Delta f_{pr} \quad (6.13a)$$

$$\sum_{g\in \mathcal{G}} \sigma_i^g + \sum_{a\in A} \sigma_i^a \geq R_{max} \quad (6.13b)$$

Constraints on generators:

$$0 \leq P_i^g \leq \varphi_i^g \cdot P_{max,i}^g \quad \forall g \quad (6.14)$$

$$P_i^g + \pi_i^g \leq \varphi_i^g \cdot P_{max,i}^g \quad \forall g \quad (6.15a)$$

$$P_i^g + \sigma_i^g \leq \varphi_i^g \cdot P_{max,i}^g \quad \forall g \quad (6.15b)$$

$$P_{i+1}^g + \pi_i^g \leq \varphi_{i+1}^g \cdot P_{max,i+1}^g \quad \forall g \quad (6.16a)$$

$$P_{i+1}^g + \sigma_i^g \leq \varphi_{i+1}^g \cdot P_{max,i+1}^g \quad \forall g \quad (6.16b)$$

$$0 \leq \pi_i^g \leq \pi(\varphi)_{max,i}^g \quad \forall g \quad (6.17a)$$

$$0 \leq \sigma_i^g \leq \sigma(\varphi)_{max,i}^g \quad \forall g \quad (6.17b)$$

$$-\rho_g \cdot \Delta t \leq P_{i+1}^g - P_i^g \leq \rho_g \cdot \Delta t \quad \forall g \quad (6.18)$$

Constraints on thermostatic loads:

$$P_{min,a} \leq P_d(\mathcal{S}^a)_i^a \leq P_{max,a} \quad \forall a \quad (6.19)$$

$$0 \leq \pi_i^a \leq P_d(\mathcal{S}^a)_i^a - P_{min,a} \quad \forall a \quad (6.20a)$$

$$0 \leq \pi_i^a \leq P_d(\mathcal{S}^a)_{i+1}^a - P_{min,a} \quad \forall a \quad (6.20b)$$

$$0 \leq \sigma_i^a \leq P_d(\mathcal{S}^a)_i^a - P_{min,a} \quad \forall a \quad (6.21a)$$

$$0 \leq \sigma_i^a \leq P_d(\mathcal{S}^a)_{i+1}^a - P_{min,a} \quad \forall a \quad (6.21b)$$

$$S^r_{min,_a} \leq S^a_i \leq S^r_{max,_a} \quad \forall a \quad (6.22)$$

$$S^a_{i+1} - \tau_a \cdot \left(1 - e^{-\frac{\Delta t}{\tau_a}}\right) \cdot \sigma^a_i \geq S_{min,_a} \quad \forall a \quad (6.23a)$$

$$S^a_{i+2} - \tau_a \cdot \left(1 - e^{-\frac{\Delta t}{\tau_a}}\right) \cdot \sigma^a_i \geq S_{min,_a} \quad \forall a \quad (6.23b)$$

$$\frac{1}{m}\sum_{t=1}^{m} S^a_i = S^a_0 \quad \forall a \quad (6.24)$$

$$S^a_1 = S^a_{48+1} \quad \forall a \quad (6.25)$$

Equations (6.12-13) are system level constraints. In particular, the pre-fault power flow F_i through the tie line is limited at each interval by constraint (6.11). Constraints (6.12) represent the power conservation equations for the two nodes of the grid. The provision of the primary and secondary response is guaranteed by (6.13). For the primary response (6.13a), the requirement at each interval is slightly lower than R_{max}, the maximum expected generation loss (the same at each time step). This is due to the transient demand damping effect on frequency, and it is introduced by

$$K_i = \frac{q}{f_0}\left\{\sum_{a\in A}[P_d(\mathcal{S}^a)^a_i - \pi^a_i] + P_{s,_i}{}^{(N)} + P_{s,_i}{}^{(S)}\right\} \quad (6.26)$$

with f_0 the nominal frequency and q the damping coefficient [58]; in addition Δf_{pr} (negative quantity in case of generation shortage) is the maximum intermediate steady state frequency deviation imposed by the TSO after the deployment of primary response (-0.5Hz [20]). The requirement (6.13b) for secondary response does equal $R_{max},$.

The second set of constraints regard the generation units. The generators' output are restricted by (6.14). Constraints (6.15) impose physical limits for primary and secondary response for the generation units. A frequency event may occur at any time within an interval, and the secondary response commitments must be provided for 30 minutes (one time interval) from that moment onward. For this reason the services agreed for interval i must also be ensured for the interval $i+1$, resulting in equations (6.16). The maximum frequency response provided by generators (6.17) is also affected by their physical limits. In particular,

$$\pi(\varphi)^g_{max,i} = \left[\left(\Delta f_n \cdot \varphi^g_i \cdot P_{max,_i}{}^g\right)/\left(f_0 \cdot \delta_g\right)\right] \quad (6.27)$$

with δ_g is the governor droop, Δf_n the maximum frequency deviation (absolute value) when frequency is at the nadir (0.8Hz [20]); for the secondary response we use

$$\sigma(\varphi)^{g}_{max,i} = \varphi_i^g \cdot {P_{max,_i}}^g \cdot \rho_g \cdot t_s \tag{6.28}$$

with the ramp rate ρ_g expressed in MW/min and t_s the time to deliver the service (30s). Constraint (6.18) limits the upwards and downwards power ramps of generators between adjacent intervals.

The last set of constraints deals with the responsive devices; the electrical power of TCLs is limited by the controller's power constraints (6.19). The same aim of (6.15-6.16) characterises the participation to security services by TCLs in (6.20-6.21). Equation (6.22) considers the TCLs energy limitations for regular operation, while (6.23a) ensures that the energy limits associated with the provision of secondary response are respected. The energy constraint is also enforced for the subsequent interval $i+1$, resulting in (6.23b). The primary response is neglected for the energy constraints as the duration of this service is limited, thus the energy decrease is negligible. Finally, constraint (6.24) implements (6.4)[8] and (6.25) imposes the recovery of the initial energy level at the end of the optimization horizon. This also avoids the 'free lunch' solution that empties the TCLs' thermal energy tank at the end of the day. Problem (6.10)-(6.25) has been solved by means of the quadratic programming routine "quadprog" in Matlab [74].

6.4 Case study

We consider the temperature models of four distinct types of refrigeration units across the domestic and commercial sectors. We represent the domestic sector using fridge-freezers, while within the commercial area we take into account bottle coolers (glass doors), freezers and multideck units (open retail units). Parameters for the TCLs dynamic models are listed in Tables 7 and 8 (Chapter 3) and they are used to compute energy (6.3) and power limits (6.6). The emergency energy intervals are assumed to be linked to the temperature bounds of the hysteresis loop of each appliance type, and the regular energy interval is 75% of this range with respect to S_0. Table 10 summarizes the number of commercial units considered, sorted by different business.

[8] See appendix to Chapter 5 for the mathematical derivation of (6.24) starting from (6.4).

We implement 45 million of domestic refrigeration units. We also consider a domestic unit in each hotel room; to this end we use an average number of rooms per hotel [83] multiplied by an average rate of room occupancy [83]. The aggregate power of TCLs across the four classes reaches 3 GW. The results illustrated in Section 6.5 represent the scenario where all appliances are responsive, but sensitivity studies with lower penetrations of TCLs are discussed in Section 6.5.1.

Table 10 Classification of commercial refrigeration units

	Pubs [84]	**Coffees** [85]	**Food Retails** [86]	**Hotels** [87]
Number	50000	6000	444050	47000
Bottle cooler	*3*	*3*	*3*	*6*
Freezer	*3*	*2*	*3*	*4*
Multideck	*/*	*3*	*2*	*/*

The GB generation mix is described by eight different power generating technologies, each representing a population of individual generators. For the population parameters used in this case study we refer the reader to Table 12 in the appendix of this chapter. Moreover, Figure 40 shows the available wind power assumed at both nodes.

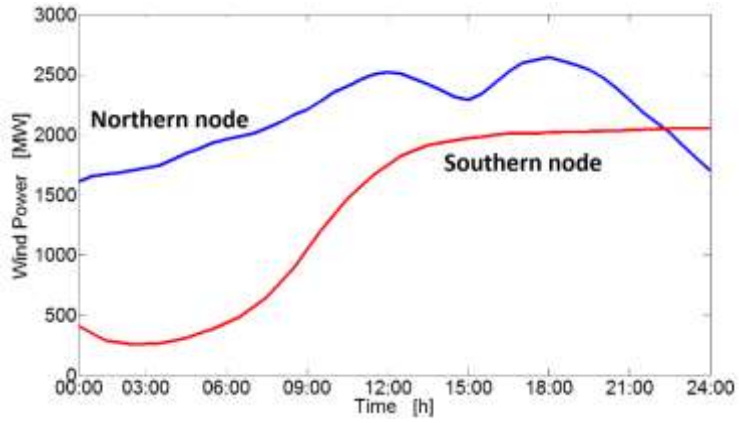


Figure 40 Maximum wind output (blue in the North, red in the South; it may be curtailed as required). Wind generation is assumed to be perfectly forecast.

The power requirement for frequency response R_{max} is set to 1.8 GW [12]; we adopt a typical value for the governor droop, 5% and 1 for the damping coefficient q [58]; in addition wind farms and nuclear generation do not provide any frequency service [46]. Furthermore, we make use of a representative winter GB demand profile [79] for which 95% of the total demand is located in

the southern node. The maximum power transfer through the tie line is set to 3300 MW [88]; further studies varying the maximum power transfer permitted are in Section 6.5.1.

6.5 Results

This section illustrates the results obtained solving the optimisation problem (6.10-6.28). We first focus on the system perspective. The system demand profile is shown in Figure 41.

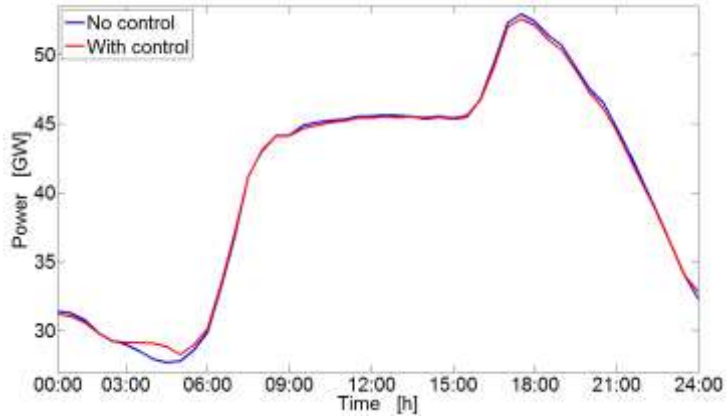


Figure 41 Impact of TCLs' flexibility on demand profile. The whole load is static (blue); the system includes TCLs flexibility (red).

The red profile refers to the case with flexible TCLs' consumption and it is compared to the reference profile (blue) that does not consider the TCLs' response. Due to the relatively small fraction of flexible demand considered compared to total system load, it is not possible to achieved a significant reduction of the demand peak. However the beneficial impact of TCLs' flexibility on the electricity cost is clearly highlighted in Figure 42A; the marginal costs at the two nodes differ when the transmission corridor is saturated. Moreover, around 17:30 the system demand peak occurs (see Figure 41) and the electricity cost at the southern node (red dashed) registers a spike because the load power peak is supplied by expensive peaking generators (Figure 42B, blue solid). These units are normally called on line only during those hours when the system demand is particularly high. With a flexible TCLs consumption instead, the cost spike significantly reduces (Figure 42A red solid) due to a reduction of the TCLs power consumption at the southern node (Figure 43B green solid). This peak-shaving effect prevents committing peaking generators (Figure 42B dotted) and thus lowers the electricity cost. In fact this demonstrates how TCLs contribute to transmission congestions' relief.

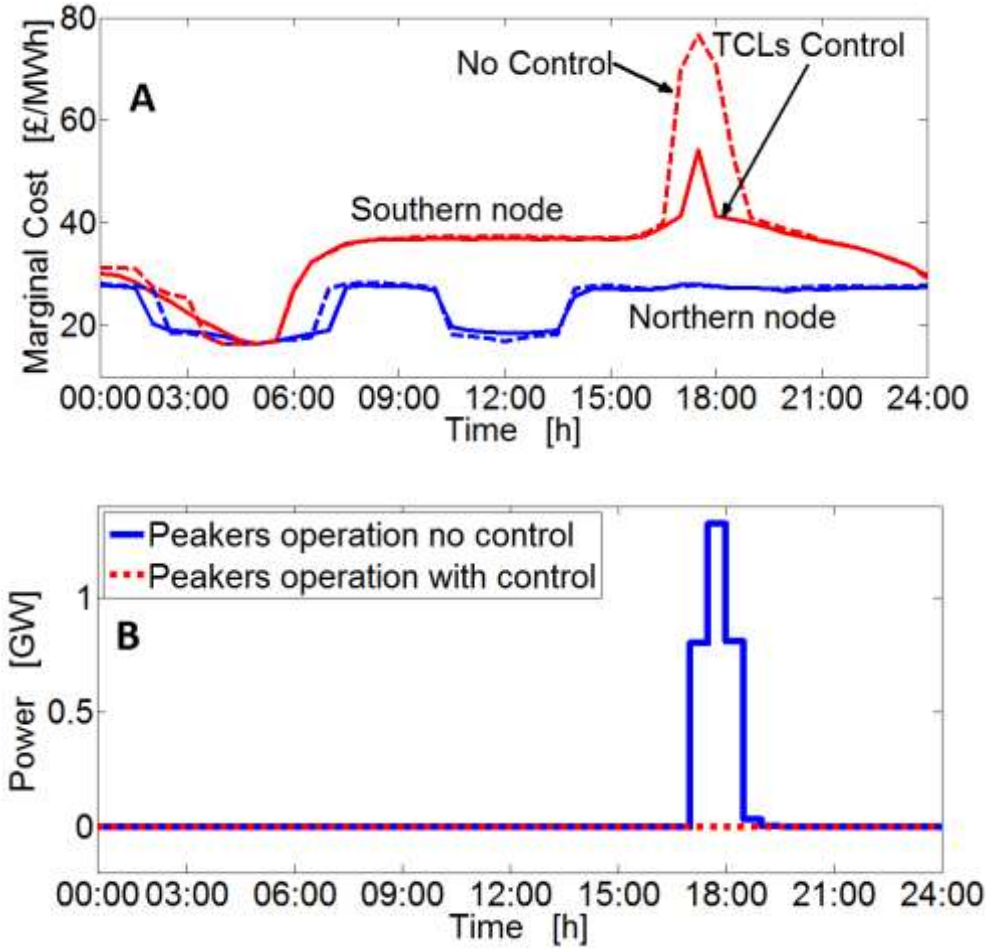


Figure 42 (A) Electricity marginal costs at northern (blue with control, dashed no control) and southern (red with control, dashed no control) node. (B) 'Peaking generators' dispatch (blue no control, red-dotted with flexible TCLs)

The associated device' behaviour for domestic refrigerators is illustrated in Figure 43; the energy (A) and power (B) profiles are shown in blue (north) and green (south). Similar profiles are obtained for the other refrigeration units. The aggregation of individual TCLs is controlled such that the instantaneous energy/power level does change during the day but it remains between precise energy and power limits (6.19, 6.22 and 6.24). As expected, the TCLs tend to increase their energy levels when (local) costs are low and decrease their energy levels when costs are high, but the cost signals are quite different in the two nodes (cheap generators are always being available in the North).

It is worth pointing out that TCLs at the southern node increase the thermal energy stored just before the system demand peak occurs; as the demand peak corresponds to high cost, these TCLs can exploit the energy reservoir to allow a power reduction. Moreover, the amplitude of this reduction is the result of a trade-off between reducing the consumption to relieve the congestion

(possibly de-committing another type of generators) and maintaining the consumption quite high to help in the fulfilment of system response requirements.

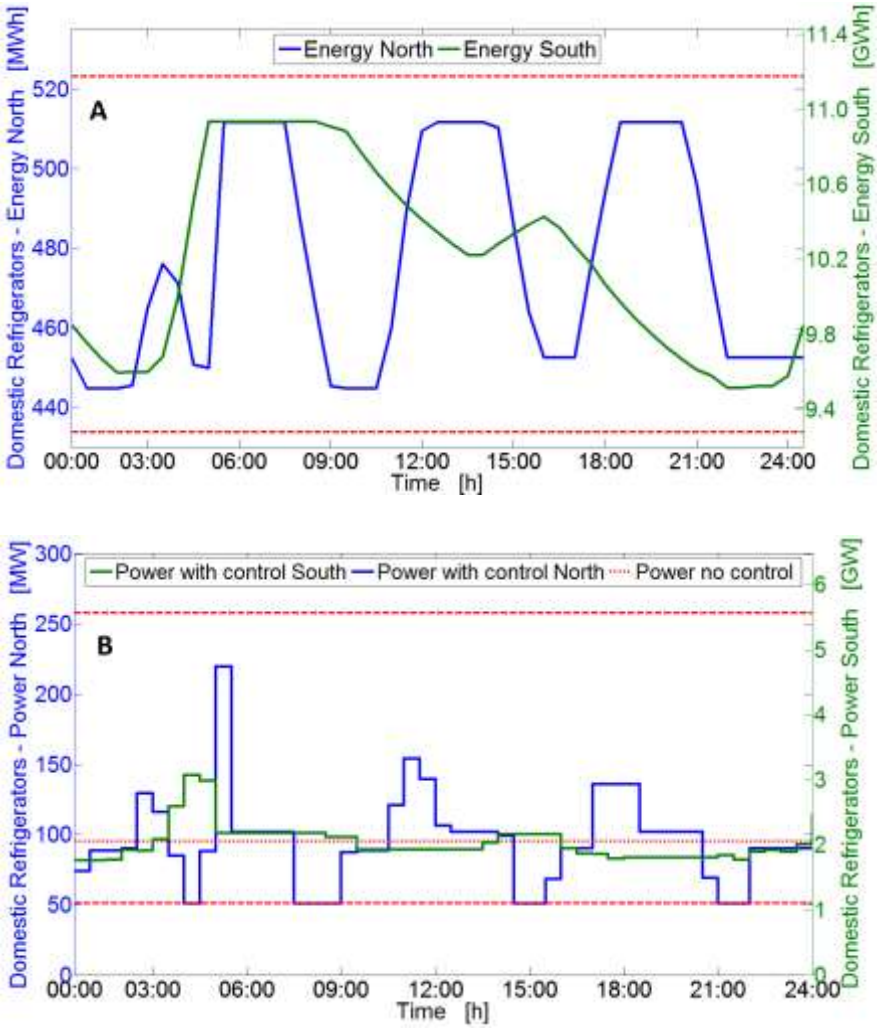


Figure 43 Energy (A) and Power (B) profiles of domestic fridge-freezers (blue for North, green for South); dashed lines illustrate the emergency energy (A), power (B) limits; regular energy interval is 75% of the emergency interval. The dotted line in (B) illustrate the emergency energy (A), power (B) limits; regular energy interval is 75% of the emergency interval. The dotted line in (B) illustrates the nominal steady state power profile without flexible control of TCLs.

Conversely, when the system peak occurs, the flexible loads in the North increase the energy and keep a power level higher than the nominal value; by doing this, these devices can help the fulfilment of the response requirement (southern units are now providing less). As the aggregate power consumption in the northern node is low and is supplied by cheap wind farms, this action does not increase the electricity cost at that node. This effect permits to make a better use of renewable generation in the northern node, otherwise limited due to constraint (6.12). The value for the system to adjust the TCLs consumption in response to network needs is illustrated in Figure 44.

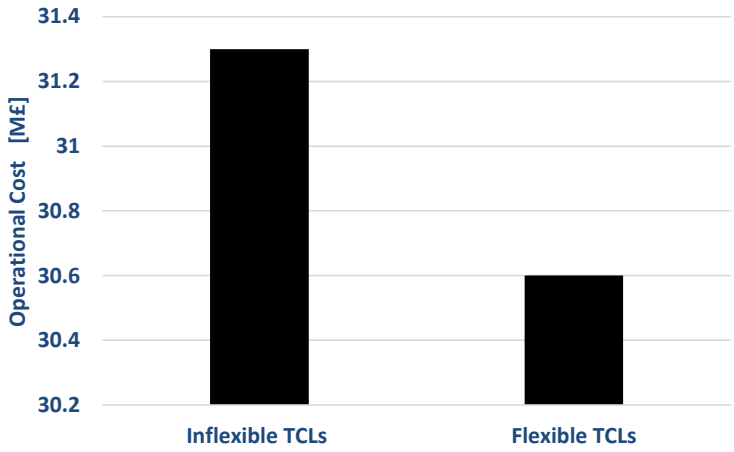


Figure 44 Daily system operational cost

Exploiting the TCLs flexibility offers 2.3% reduction of the system operational costs.

6.5.1 Sensitivity studies

Three case studies reflecting the National Grid operation of the Anglo-Scottish boundary are now considered; in the first, the power transfer is set to 3300 MW [88]. Future reinforcements will increase this value up to 4400 MW [88] (second case); in the last case the power flow is left unconstrained. We also consider different penetration levels of smart appliances, ranging from 0-100%. Figure 45 shows the sensitivity of the system operation cost with respect to these variations, expressed as the difference of operational costs with and without (reference case) controllable thermostatic units.

The cost savings increase with the size of the active TCLs population; fixing the TCLs' penetration, the savings slightly decrease if the system allows a higher power transfer through the tie line. Additional capacity allows cheaper generation in the North to replace some expensive units in the South even without responsive TCLs.

The device perspective is considered in Figure 46, which illustrates the daily savings obtained for each individual TCL as a function of the overall TCLs penetration in the system. As higher savings are achieved for a low penetration of TCLs, it rises the incentive for entities like aggregators or single customers to start and spread the implementation of flexible TCLs for demand response mechanisms. In particular, the limit case of 0% penetration corresponds to the savings achieved by the first marginal TCL units integrated in the system.

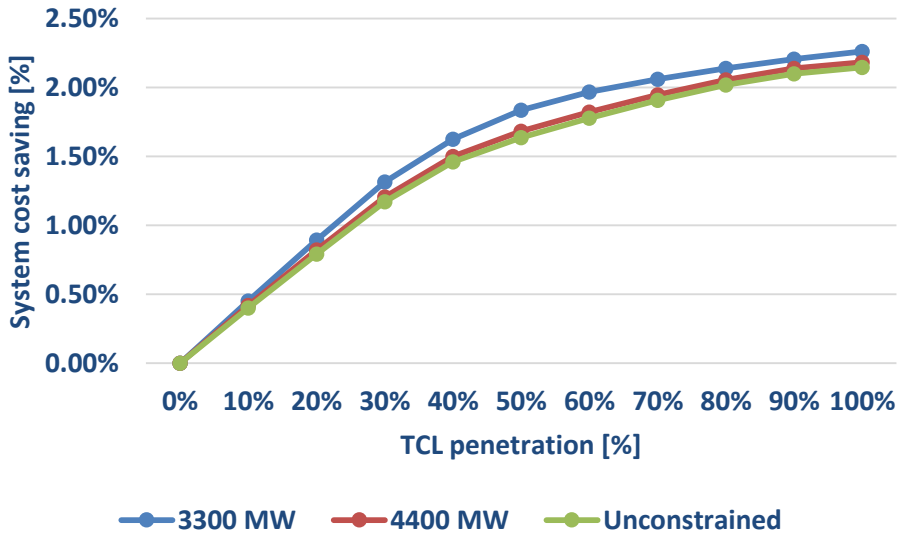


Figure 45 Sensitivity of system cost percent savings to the TCLs penetration for three power transfer (3300 MW blue, 4400 MW red, unconstrained green)

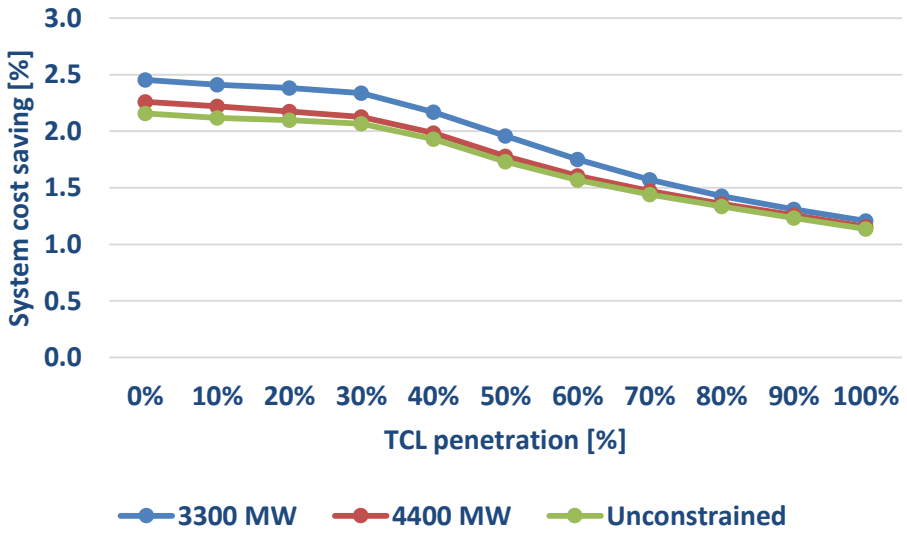


Figure 46 Daily savings for individual TCL varying the TCLs penetration for three maximum power transfer (3300 MW blue, 4400 MW red, unconstrained green)

Finally in Table 11 we analyse the CO_2 emission savings due to the TCLs support; in particular, varying the power flow limit of the tie line, we assess the trend of the emission savings compared to the trend of the system cost savings. Note that the CO_2 emissions were not part of the objective function, so these savings are a by-product of the cost reductions attained[9]. The impact of TCLs on CO_2 emissions increases as the transmission constraints are reduced. In the highly constrained scenario (F_{max}=3300 MW) TCLs drastically reduce the operational cost as they avoid the dispatch of

[9] In Chapter 7 the wind curtailment is directly penalised in the objective function

expensive peaking generators (high cost savings); however, due to the low power flow limit, the contribution of wind and nuclear generation (low carbon footprint) in the northern node is still limited.

Table 11 Daily system cost and co2 emission savings

	System Cost [M£]			**CO_2 Emissions [Mtons]**		
	3300 MW	**4400 MW**	**No limit**	**3300 MW**	**4400 MW**	**No limit**
TCLs 0%	31.3	31.1	31.0	0.503	0.496	0.496
TCLs 100%	30.6	30.3	30.3	0.501	0.490	0.489
Percent variation	-2.3%	-2.18%	-2.15%	-0.56%	-1.19%	-1.28%

Peakers generators instead are replaced by other conventional sources in the southern node, which are cheaper but still have a high carbon footprint. In the unconstrained scenario the opposite situation takes place; the TCLs' flexibility permits to integrate even more wind and nuclear generation from the northern node realising a high CO_2 emission curtailment. However, the percent system cost reduction is low as peakers generators were not required even in the reference scenario (0% TCLs penetration – unconstrained flow).

6.6 Conclusions

In this chapter we have quantified the value for the security constrained economic dispatch to integrate flexible TCLs aggregate consumption in order to minimise the system operational cost. Provided the ability of thermostatic loads to shift their energy consumption, we recognized the beneficial impact that this action has on the electricity cost spike reduction in a congested power system. As result of an aggregate power reduction, thermostatic appliances permit to turn off high cost peaking generators that are normally called up only in order to supply the system demand when it achieves the highest values. In this sense, TCLs can temporarily alleviate transmission lines congestions. Moreover, we illustrated how the optimal charging/discharging policy for TCLs may considerably differ if the devices are connected to different bus bars of the system.

Further benefits due TCLs flexibility are encountered when the responsive devices support the provision of frequency services. This action diminishes the need to run generators part-loaded; as consequence of this, the power system is operated more efficiently, hence with a lower operational cost. The overall cost reduction enabled by TCLs also results in a consequent reduction of CO_2 emissions as TCLs can make up for the lack of flexibility of non-emitting generators. Finally, we also quantified the daily savings obtained for each individual TCL as a function of the overall TCLs penetration in the system; higher savings are achieved for a low penetration of TCLs and this may represent a strong incentive for entities like aggregators or single customers to start and spread the implementation of flexible TCLs for demand response mechanisms.

6.6.1 Extensions and links with the next chapters

The results obtained in this chapter represent a starting point in order to quantify the benefits associated with the integration of demand side response mechanisms in the network operation. However, the analysis carried out so far did not consider a number of aspects regarding the power system model and demand side response features. In particular, it would be required to assess the impact that the simultaneous provision of services will have on the commitment of individual generation in presence of larger penetration renewable energy sources. Particular attention should be dedicated on the uncertainty that characterizes wind generation and therefore it arises the need for stochastic scheduling routine that properly recognizes the uncertainty of wind. Moreover, the calculation of the system frequency requirements should reflect the strong impact that the future lack of system inertia will have on the maximum RoCoF and the minimum value of frequency after a generation outage.

In addition, if the system frequency requirements will depend on the actual level of inertia, they may be considerably different during the whole day depending on the actual net demand i.e. system demand minus wind generation available at each step. An accurate assessment of the demand side response value cannot exclude the effect of the load recovery after the provision of response services. In fact, the extra power consumption absorbed by TCLs to recover a determined energy (temperature) level has to be included in the calculation of reserve requirements; this, in turn, will impact the overall system operational cost and represents a key ingredient to optimally allocate TCLs response services. In the next chapter, we increase the complexity of the problem taking into consideration the improvements listed above.

Appendix

The data of the generators' costs (β and γ) and the maximum capacity and are courtesy of Dimitrios Papadaskalopoulos; fixed costs α are courtesy of Fei Teng. Data of ramp rates are from [89] and CO_2 emission coefficients are from [90].

Table 12 Generation units characteristics

Technology	Capacity [GW]		Fixed Cost α [£/h]	Linear Cost β [£/MWh]	Quadratic Cost γ [£/(MWh)^2]		Ramp Rate [%P_{online}/min]	CO_2 Emissions [gCO2/KWh]
	North	South			North	South		
Wind	3.87	3.35	0	0	0	0	n.a.	5
Nuclear	1.47	4.61	300	0	0	0	0	5
Base Gas	0.38	10.8	4000	24	0.010	0.0004	8.3	500
Base Coal	1.10	8.82	2500	28	0.004	0.0005	2	1000
CHP-Biomass	0.23	1.90	2000	48	0.032	0.0061	4	100
Hydro	0.72	/	0	54	0.014	/	100	10
Marginal Gas	0.38	10.2	5000	62	0.030	0.0011	10	500
Marginal Coal	1.10	8.18	3000	70	0.011	0.0016	4	1000
Pump Storage	0.47	0.64	100	100	0.054	0.0200	50	5
Peakers	/	3.35	17000	130	/	0.0142	8.3	650

Chapter 7: Stochastic Scheduling with Flexible Thermostatically Controlled Loads

Abstract

The benefits to power system operation and security stemming from the TCLs' provision of frequency services and energy arbitrage have been investigated in previous studies. However, these system services were considered individually. Furthermore, the modelling and control of the energy payback consequent to the actual provision of these services have often relied on approximations or, in some cases, control methods have been designed to *a priori* cancel any extra power absorption; this way the potential flexibility of demand side response reduced. In this chapter, an upgrade of the DSRM introduced Chapters 5 and 6 is proposed; the improvements achieved now permit to accurately consider the load energy recovery dynamics. It allows for the optimal allocation of frequency services in accordance with the amount of extra power to be absorbed during the payback period. The cost of supplying this extra power is automatically taken into account in the optimal allocation of TCLs frequency response. The proposed framework, integrated in a SUC routine, permits to achieve the aim of this thesis and thus quantifying the value of demand side response with TCLs.

Nomenclature

The main variables and parameters used in this chapter are given below; others are provided throughout the chapter.

Variable	Description	Unit
$\boldsymbol{n}$	*Node of the scenario tree*	
$\boldsymbol{a(n)}$	*Parent node of node n*	
$\boldsymbol{\Delta\tau(n)}$	*Time interval corresponding to node n*	h
$\boldsymbol{\pi(n)}$	*Probability of reaching node n*	
$\boldsymbol{\mathcal{G}}$	*Set of conventional generators*	
$\boldsymbol{g}$	*Subscript for conventional generators*	
$\boldsymbol{C^{LS}}$	*Value of lost load*	£/MWh
$\boldsymbol{P_g^{max}}$	*Maximum production of generator g*	MW
$\boldsymbol{R_g^{P,max}}$	*Maximum primary response capability of generator g*	MW
$\boldsymbol{R_g^{S,max}}$	*Maximum secondary response capability of generator g*	MW
$\boldsymbol{r_g^{max}}$	*Maximum reserve ramp rate of generator g*	MW
$\boldsymbol{f_g^p}$	*Proportion of spinning headroom for primary response*	
$\boldsymbol{f_g^S}$	*Proportion of spinning headroom for secondary response*	
$\boldsymbol{C_g(n)}$	*Operating cost for generator g*	£
$\boldsymbol{P_{LS}(n)}$	*Load shed at node n*	MW
$\boldsymbol{P_W(n)}$	*Realized available wind power at node n*	MW
$\boldsymbol{P_{WC}(n)}$	*Wind curtailment at node n*	MW
$\boldsymbol{P_D(n)}$	*Total static demand at node n*	MW
$\boldsymbol{P_T(n)}$	*TCLs power consumption at node n*	MW
$\boldsymbol{P_g(n)}$	*Power output of generator g at node n*	MW
$\boldsymbol{R_g^P(n)}$	*Primary response provision from generator g at node n*	MW
$\boldsymbol{R_g^S(n)}$	*Secondary response provision from generator g at node n*	MW
$\boldsymbol{P_T^P(n)}$	*Primary response provision from TCLs at node n*	MW
$\boldsymbol{P_T^S(n)}$	*Secondary response provision from TCLs at node n*	MW
$\boldsymbol{R_g^R(n)}$	*Reserve provision from generator g at node n*	MW
$\boldsymbol{P^P(n)}$	*Primary response requirement at node n*	MW
$\boldsymbol{P^S(n)}$	*Secondary response requirement at node n*	MW
$\boldsymbol{P^R(n)}$	*Reserve requirement at node n*	MW

$\boldsymbol{S_T(n)}$	*Energy level of TCLs at node n*	MWh
$\boldsymbol{t_p}$	*Delivery time for primary response*	s
$\boldsymbol{t_r}$	*Delivery time for reserve*	s
$\boldsymbol{N_g^{up}(n)}$	*Operation status (0/1) of generator g*	
$\boldsymbol{\Delta f_{max}}$	*Maximum frequency deviation*	Hz
$\boldsymbol{\Delta f_{max}^{iss}}$	*Maximum frequency deviation at intermediate steady state*	Hz
$\boldsymbol{RoCoF_{max}}$	*Maximum RoCoF requirements*	Hz/s

7.1 Introduction

The integration of large shares of renewable energy sources (RES) aims to diminish greenhouse gas emissions. However, including these generation technologies in the system operation will require an increased amount of various ancillary services due to the intrinsic variability and uncertainty of RES and their limited ability to support the system inertial-frequency response. In particular, the lack of system inertia exacerbates the need for fast frequency regulation services in order to maintain the initial frequency drop within security boundaries and avoid costly emergency demand disconnections [91] or, in the worst case, a black out. Following a traditional approach, additional ancillary services would be still delivered through an increased amount of part-loaded generators in combination with fast standing and high-cost plants. This not only decreases the system efficiency and leads to higher operation cost, but it may also compromise the final ability of the system to integrate RES.

An alternative approach suggests that the fulfilment of these system requirements may be facilitated by demand side contribution. Initial research [43, 44, 45] has investigated the value for DSR providing reserve in joint energy/reserve markets. However, the characteristics of the DSR models in these works did not match with the physical constraints of any actual demand side technology (e.g. battery storage, electrical vehicles or thermostatic loads). In this thesis, we focus on thermostatic loads that, within certain boundaries, are generally not time-critical and can sustain small alterations to the regular duty cycle.

The authors in [46] quantified the value for system scheduling if TCLs provide primary control. In addition, these appliances may enhance the system operation by making energy

arbitrage [34, 47]. Similar studies, though based on different technologies such as battery storage and electrical vehicles, revealed the benefits of selecting optimal portfolios of services to be delivered simultaneously rather than individually [26]. However, as pointed out in Chapter 5, achieving appropriate level of coordination for TLCs is not straightforward: in fact, individual appliances typically have only an on and off power states [49] and consequently the control frameworks in [46, 47] enable the provision of individual services.

In Chapters 5 we introduced a platform for the simultaneous provision of multiple services, such as energy arbitrage and frequency response, using TCLs. Although this platform allows for accurate control over both short term and long term time scales, its application to the SCED problem (Chapter 6) did not fully considered the effect TCLs´ energy recovery on the system operation after the provision of secondary response. However, a reliable supply of demand side services has to be guaranteed regardless of the provision of the same service in previous intervals. Hence, an accurate assessment of the demand side response value cannot neglect the effect of load recovery and its associated cost.

The energy recovery could be performed by means of extra power consumption after the deployment of the response services. In [45, 92] this power absorption is approximately modelled as a function of the previous power reduction through a generic constant parameter. Instead, the extra consumption was suppressed in [46] by means of an *ad hoc* control strategy. Nonetheless the ability to suppress the payback is paid at the cost of a considerably slow energy recovery; this design automatically prevents TCLs from supplying medium term response services, limiting the contribution of TCLs only to very fast response services.

In this chapter, an advanced version of the DSRM discussed in the previous chapters is presented; it accurately includes the dynamics of the load recovery pattern and calculates the associated post fault energy levels. This novel model, in combination with a system scheduling routine, permits to calculate the optimal allocation of flexible energy consumption and response service provision of a heterogeneous population of TCLs that minimizes the system operational cost. The optimal solution considers the actual cost of extra generation reserve to assist the load recovery.

The provision of frequency response can vary at each time step in accordance with the time dependent characteristic of the system requirements. These requirements depend on the level of inertia that, in turn, reflects the high variability of wind. The proposed DSRM methodology is

constructed in such a way that TCLs would always guarantee the deliverability of the scheduled response services as the energy deployed is fully paid pack by the end each time interval. This characteristic makes the supply of DSR services highly reliable and comparable with generators' standards. In addition, the feasibility of the TCLs´ energy profile is guaranteed by means of the accurate decentralised control described in Chapter 4.

The scheduling routine used is a mixed integer linear programming (MILP) formulation for the SUC. The proposed scheduling model refers to a joint paper with Fei Teng [57]. It is worth pointing out that a common deterministic unit commitment would have shown the expected benefits associated with the DSRM as well, with regard either to the optimal TCL energy scheduling either to the provision of ancillary services (primary and secondary response) by controlled appliances. In effect, at present, ancillary services are scheduled by imposing deterministic pre-defined requirements in the generation scheduling procedure. However, the aim of this chapter is to evaluate the benefits of using TCL support in a future power system characterized by large shares of wind generation; as the uncertainty introduced by wind production is more significant than that affecting the system demand, a scheduling process performed under such deterministic rules may be inefficient. Hence, in order to increase the accuracy of the results obtained in this chapter we decide to adopt a stochastic scheduling routine for the analysis carried out.

The author of this thesis has been focused on the mathematical formulation of novel frequency constraints. Hence, in this chapter the SUC framework will be briefly recalled and more attention is paid to the mathematical formulation of the inertia-dependent frequency constraints. Specifically, three novel constraints ensure that the dynamic evolution of post-fault frequency will satisfy security requirements associated with the maximum RoCoF, the frequency nadir and the intermediate steady state frequency limit, with respect to the GB security standards [20]. It is worth pointing out that the first two constraints depend on the system level of inertia. Moreover, the proposed model directly incorporates the requirement for system dynamic frequency performance into stochastic scheduling framework so that the impact of wind uncertainty on the system inertia can be recognized.

Finally, note that the SUC considers the TCL energy consumption and their aggregate operation is ensured by the hybrid (stochastic-deterministic) control strategy discussed in Chapter 4. If the stochasticity of the scheduling routine in this chapter is effectively linked to the uncertainty of wind production, the one regarding the controller in Chapter 4 is related to the random status

jumps that each TCL has to perform in order to follow a pre-determined power profile. It follows that there are no similarities between these two stochastic models. Moreover, as we discussed in Chapters 4 and 5, in the limit of a large number of devices, the aggregate TCL power response can be treated as a deterministic response. In effect, the leaky storage model in Chapter 5 is the one integrated in the SUC and it can be considered as a deterministic model under the assumptions previously provided. With more detail, the stochasticity in each individual TCL behaviour and wind occurs on very different time scales i.e. few seconds the former, several minutes the latter. For this reason these can be really treated separately and do not influence each other.

The rest of the chapter is organized as follows. Section 7.2 offers a detailed description of the DSRM characteristics alongside of the mathematical formulation of the TCLs´ model. Section 7.3 deals with the stochastic unit commitment highlighting the modifications made to the original framework in [57] in order to integrate the constraints associated with the demand response. Furthermore, in Section 7.4 the modelled system response and reserve requirements for the services are formulated, whereas in Section 7.5 the case study and results are presented. In particular, we compare our methodology with other possible solutions to show the effectiveness of the proposed work. In addition, discussions regarding the optimal recovery pattern and the impact of different implementations of the average energy constraint complete the section. Finally, Section 7.6 concludes the chapter and suggests further improvements to this work. The Appendix of this chapter is located after the conclusions.

7.2 Demand Side Response Model: main features and mathematical formulation

In Section 7.2.1 we first recall the formulation introduced in Chapter 5 to describe a heterogeneous cluster of thermostatically controlled load as a leaky storage unit with envelope constraints. In Section 7.2.2 we explain the main features of the proposed DSRM; the mathematical formulation of the model follows in Section 7.2.3. Finally, the comparison against alternative methods concludes this section.

7.2.1 Aggregate heterogeneous TCLs

A large heterogeneous population of TCLs can be described as a leaky storage unit with associated envelope constraints. The envelope parameters, labelled with the hat, do not represent

a particular "real" device within the cluster; they just bound the flexibility of the whole TCLs' population. Hence, the energy evolution ($S(t)$ [MWh]) of a cluster of $N \gg 1$ different TCLs is described by:

$$\frac{dS(t)}{dt} = -\frac{1}{\hat{\tau}}S(t) + P(t) \tag{7.1}$$

with $\hat{\tau} = \min_a \tau^a$ the thermal time constant [h]; the superscript a is used for appliance-specific parameters. Moreover $P(t) \equiv \hat{P}_0 \Pi(t)$ [MW] is the power consumed; $\hat{P}_0 = \sum_a P_0^a$ [MW] is the aggregate steady state power consumption and $\Pi(t)$ is a relative power curve ($\Pi_0 = 1$ for a steady state condition). In addition, quality of service issues on individual appliances impose energy bounds on the aggregate capacity.

$$\hat{S}_{min} = \max_a S_{min}^a \leq S(t) \leq \hat{S}_{max} = \min_a S_{max}^a \tag{7.2}$$

Energy bounds alone cannot adequately respect the primary function of the TCL (cooling/heating) as the devices would be stuck at all times at the lower or upper energy bounds. Therefore, we require the mean value of the energy across a time window of interest w not to deviate too much from $\hat{S}_0 = \hat{P}_0 \hat{\tau}$, i.e. the steady state energy level.

$$\frac{1}{w}\int_w S(t)\mathrm{dt} \approx \hat{S}_0 \tag{7.3}$$

The population of TCLs is controlled by means of the two-stage control method discussed in Chapter 4. This methodology enables individual TCLs to track a relative power curve $\Pi(t)$ so that the aggregate power consumption targets such a profile in expectation:

$$E[P(t)] \cong \hat{P}_0 \Pi(t) = P(t) \tag{7.4}$$

The controller implementation introduces limits on the accessible range of power consumption levels. This implicitly defines the minimum and maximum power limits (see equations (6.7) and (6.8)).

$$\hat{P}_{min} = \max_a P_{min}^a \leq P(t) \leq \hat{P}_{max} = \min_a P_{max}^a \tag{7.5}$$

The respect of constraints (7.2) and (7.5) is sufficient to guarantee the feasibility of the response, with no need for device-level simulations. With this strategy, TCLs can follow a power profile $P(t)$ and simultaneously deliver response services, so long as their simultaneous provision does not violate appliances' constraints.

7.2.2 Demand Side Response Model characteristics

The DSRM introduced in this chapter exhibits three main characteristics, the first of which already was already included in model developed in the previous chapters.

- ***Flexible response provision***: In the DSRM the TCLs energy level at the beginning and the end of each interval (and thus the consequent power consumption) is variable; this characteristic enables a flexible provision of response services[10] in accordance with the time-dependent system requirements [57]. In fact, during hours characterized by low net demand (system demand minus the wind production) the system response requirements would be typically high due to an overall shortage of inertia. A growth in the power absorption allows for a larger provision of frequency response services. It is worth pointing out that this behaviour is in synergy with the aim of energy arbitrage as, under these system conditions, the energy cost would be typically low, facilitating the increase of the energy level. With high net-demand instead, TCLs would tend to reduce the power consumption and thus lowering the available response buffer; this action reflects the lower system response requirements during those hours[11]. Again this action aligns with the arbitrage aim as TCLs would be consequently scheduled to facilitate the system demand shaving, due to temporally high costs.

- ***Accurate energy recovery****:* The DSRM presented in this chapter allows for full controllability of the payback phase; in fact, after the provision of secondary response, TCLs can actually absorb an extra amount of power, compared to the regular power consumption scheduled for that interval. The amplitude of this power peak is precisely calculated based on the amount of the secondary response previously delivered; in particular, it does depend on the duration of the secondary response service and the available energy recovery window. Due to this extra power absorption, the TCLs´ energy recovery becomes much faster, flexible and

[10] Note that the response services are no longer modelled as a 24 hours commitment as in Chapter 5; in general, the response provided at each time step can vary.

[11] In this case, many conventional generators would be already on line to meet a high level of system demand; it follows that the level of inertia and the overall system governor response would be significant. This makes the response system requirements lower and thus the need for TCLs support in the fulfilment of these requirements decrease.

controllable. Furthermore, the extra consumption is drained during the reserve time window and therefore supplied by reserve generators; hence, from the system point of view the reserve requirement will have to rise. The cost of this increment cannot be neglected in a precise assessment of the demand side value.

We point out that the provision of secondary control by TCLs cannot be seen as an autonomous option to the secondary response supplied by generators; it is only able to postpone the generators power supply from a time window, when this supply would be very expensive[12], to another one, during which delivering the same amount of power would be cheaper. Hence, the use of TCLs to provide medium-term response (with the consequent load recovery) could be seen as a way to arbitrage between generators' response requirement, which is in itself an expensive service, and reserve requirement, which is cheaper, by decreasing the former and increasing the latter.

***Energy profile and security services reliability**:* Although the probability of having a generator outage is generally very low, the probability of having an outage at each step is independent from the actual outage realization at previous steps. In case of a failure at step k, if the TCLs´ recovery phase does not terminate by the beginning of interval $k+1$, it will not be possible to absorb the scheduled power consumption and to provide the scheduled response, without the risk of violating temperature device constraints. It follows that the ability to secure the system is not guaranteed, in which case, additional generation capacity would need to be engaged, thus entailing extra cost. The DSRM is based on the premise that the energy deployed while providing frequency services at the generic time interval k has to be fully paid back by the end of the same interval. This implies that at the beginning of interval $k+1$, the TCLs´ energy level will equal the regular energy level scheduled for the 'normal operation'.

The third DSRM characteristic makes the TCLs´ level of service really comparable to generators, which are normally ready to provide response shortly after reserve

[12] The rapidity in the generators' power provision is limited by technical constraints [58]. In a short time window (response) a certain amount of power would therefore require two generators to be delivered, whereas, if the time constraints are relaxed (reserve), it could be supplied only by one machine.

providers have taken over the balancing from response providers. This characteristic makes the demand side supply of security services highly reliable.

7.2.3 Mathematical formulation of the DSRM

Let us solve (7.1) at generic step k across the interval [0,t] of length Δt; we alternatively impose a constant power $P(t) = P_k$ or a linear power profile $P(t) = \rho_k t + q_k$. The two generic solutions below (7.6a-b) will be used to construct the DSRM.

$$S_k = S_{k-1} \cdot e^{-\frac{\Delta t}{\hat{\tau}}} + \hat{\tau} P_k \cdot \left(1 - e^{-\frac{\Delta t}{\hat{\tau}}}\right) \tag{7.6a}$$

$$S_k = S_{k-1} e^{-\frac{\Delta t}{\hat{\tau}}} + \hat{\tau}\rho_k \Delta t + \hat{\tau}\left(1 - e^{-\frac{\Delta t}{\hat{\tau}}}\right)(q_k - \hat{\tau}\rho_k) \tag{7.6b}$$

where S_{k-1} and S_k are the energy levels at the beginning and end of interval k, respectively. Considering the generic solutions (7.6a-b), we construct the DSRM (see Figure 47) that governs the energy consumption and the frequency response provision (with consequent payback) of TCLs at all the steps i of length Δt_1 of the optimization horizon. This interval is divided into three further sub-intervals of length $\Delta t_2, \Delta t_3, \Delta t_4$, respectively.

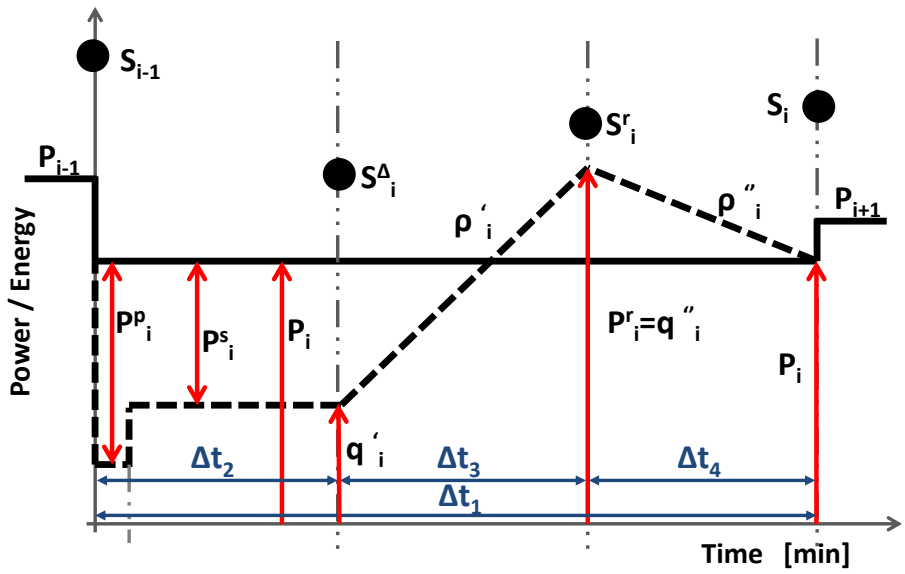


Figure 47 Schematics of the multiple service model for demand response

The initial and final energy levels are S_{i-1} and S_i [MWh] and the amount of power actually absorbed by TCLs during the interval i is equal to the constant value P_i [MW]. These quantities are limited by (7.2) and (7.5). The included services are primary (P_i^p $[MW]$) and secondary (P_i^s $[MW]$)

response. After the full deployment of primary response, the total TCLs consumption cannot drop below $\widehat{\mathrm{P}}_{min}$; adequate reserve for primary response allocation is enforced by (7.7).

$$0 \leq P_i^p \leq P_i - \hat{P}_{min} \tag{7.7}$$

For this short-term service, with a 30 s duration, the provision of primary response is assumed to not have sufficient impact to violate energy limits. The same minimum power requirement is applied to secondary response:

$$0 \leq P_i^s \leq P_i - \hat{P}_{min} \tag{7.8}$$

The power decrease for secondary response is sustained for Δt_2. Equation (7.9) ensures that the energy level at the end of the secondary control provision will not drop below the lower energy bound. Therefore, the energy storage level S_i^{Δ} respects:

$$S_i^{\Delta} = \gamma_2 S_{i-1} + \beta_2 P_i - \beta_2 P_i^s \geq \widehat{\mathrm{S}}_{min} \tag{7.9}$$

where β_2 and γ_2 are constant values. The mathematical derivation of equations (7.9-7.11) and the expressions for the constant values introduced are included in Section APP7.1 in the appendix to this chapter. Afterwards, the energy recovery phase starts and, thus, the power consumption increases with a fixed slope ρ_i' $[\frac{MW}{min}]$ from the intercept q_i' $[MW]$. In the end of the second time interval, of Δt_3duration, when secondary response has been delivered, power consumption P_i^r is greater than P_i; clearly, P_i^r and S_i^r cannot exceed the maximum power and energy level, respectively;

$$P_i^r = L_1 S_{i-1} + L_2 P_i + L_3 P_i^s \leq \widehat{\mathrm{P}}_{max} \tag{7.10}$$

$$S_i^r = H_1 S_{i-1} + H_2 P_i + H_3 P_i^s \leq \widehat{\mathrm{S}}_{min} \tag{7.11}$$

Note that the values of L_1 and L_2 are proved to be 0 and 1, respectively (see Section APP7.2); this implies that the additional amount of power P_i^{ar}, supplied by reserve generators is a proportion of the secondary response allocated by means of L_3.

$$P_i^{ar} = P_i^r - P_i = (L_2 - 1)P_i + L_3 P_s = L_3 P_s \tag{7.12}$$

In particular, considering a time step Δt_1 and fixing the secondary response commitment (Δt_2), L_3 and hence P_i^{ar}only vary with Δt_3 ($\Delta t_3 + \Delta t_4 = \Delta t_1 - \Delta t_2 = const$). A fast recovery, i.e. with a small Δt_3, leads to small P_i^{ar}, although it increases the rapidity in the power provision from reserve generators. On the other side, a large Δt_3 drastically increases the amount of reserve to be

supplied. A case study in Section 7.4.4 illustrates the impact of this trade off on the system operational cost. Finally, note that this result is in step with previous works [45, 92] However, in [45, 92] the relation between the power reduction and the consequent power to pay back was only expressed through a generic constant parameter that was empirically estimated, whereas, in this work, this relation is mathematically derived.

7.2.4 Comparison against alternative methods

We now proceed to analyse the DSRM against alternative approaches; in particular, we focus on the methodology adopted in [46] as this work deals with the economic benefits for thermostatically controlled loads providing frequency regulation. In this framework, the energy and, thus, the power consumption of the population of thermostatic devices cannot deviate from the steady state level, i.e. $S_i = \hat{S}_0$ and $P_i = \hat{P}_0 \quad \forall i$. This choice prevents the TCLs from increasing the energy/power level to provide more response services (in case of high system requirements) and, conversely, the appliances are not able to reduce the energy/power consumption to provide less response services (lower requirements) or driven by temporary high costs. Therefore, TCLs would always maintain an energy buffer ($\hat{S}_0 - \hat{S}_{min}$) that allows for a maximum response contribution limited to ($\hat{P}_0 - \hat{P}_{min}$).

Another characteristic of this approach is the absence of any extra power absorption during the energy recovery phase. In practice, a sophisticated control strategy [31] forces the aggregate devices' consumption to always reach $\hat{P}_0$ after the deployment of frequency services in case of a generator outage; from Figure 47, it results that $\left(q_i'' = P_i = \hat{P}_0, \ \forall i\right)$. The absence of any additional power consumption ($P_i^{ar} = 0$) comes at the cost of longer times for temperature recovery [31], governed by the time constant $\hat{\tau}$; in particular, for a first order model like (7.1), it would take around 4-5 times the time constant $\hat{\tau}$ to naturally reach the steady state condition $S(t) = \hat{S}_0$. Considering domestic fridge-freezers, this time would be around 18 hours ($\hat{\tau} = 4.51\ h$ see Section 7.5) and thus not acceptable, given the possibility of another generation outage during the recovery phase.

However, if TCLs (domestic fridge-freezers) provide the maximum secondary response contribution ($\hat{P}_0 - \hat{P}_{min}$), and afterwards they recover $\hat{P}_0$, the percent temperature deviation after six hours would be below 1% (see Figure 48).

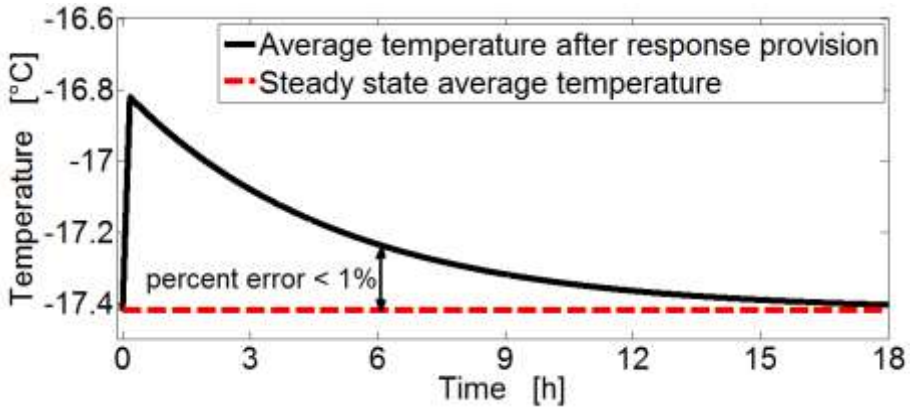


Figure 48 Average temperature evolution for domestic fridge-freezers providing the maximum secondary response contribution and recovering the steady state power level after the deployment of the service.

Therefore, we assume that after this time the TCLs population has effectively recovered the steady state condition. It follows that, after a generator failure, TCLs would not be able to provide again the response scheduled without the risk of violating temperature constraints for the next 5.5 hours. The authors in [46] decided to use only half of the TCLs population at each time step so that, in presence of a second failure, the second half of TCLs could still assist the frequency control. However, this choice would not guarantee the ability to provide response services during the next time steps.

However, in order to compare the performances of this methodology with the ones offered by our DSRM, the same level of reliability in the provision of the scheduled response services needs to be guaranteed. Hence, considering 30 minutes time steps, we impose to use only a fraction, equal to 1/12, of TCLs providing secondary response (up to the maximum power reduction) at each time step[13]; this way, TCLs would still be able to guarantee the provision of the scheduled services regardless of the actual occurrence of generators' failures at previous steps. This limitation is not applied to primary response as the energy deployment is negligible.

7.3 The stochastic scheduling model

In this section, we introduce the multi-stage stochastic scheduling model with rolling planning based on the model developed in [57]. This routine optimizes the system operation by

[13] This is assumed to be done as in [46] by means of random number generator in each simulated appliance. Alter the provision of primary response, a TCL decides whether it is going to provide or not secondary response.

simultaneously scheduling energy production, inertia-dependent primary response, load-dependent secondary response and contingency reserve, in the light of uncertainties associated with renewable production and generation outages. The unit commitment is solved over a scenario tree; a typical tree is shown in Figure 48. The scenarios are weighted according to their probability to realize,

Figure 49 Schematic of a typical scenario tree

The model optimally balances the cost of the committed generation in each scenario against the expected cost of not meeting demand and it schedules the system on a 24 hours basis over one year time horizon. Moreover, because of the rolling planning, at the second time step, the model provides a new optimal solution for the next 24 hours, based on a better forecast compared to the one had at the previous step. This process is repeated at each time step until the last interval of the year.

Quantile-based scenario selection method proposed in [57] is applied to the SUC considered. This method constructs and weights scenario trees based on user-defined quantiles of the distribution of uncertain variables. Compared with commonly used Monte Carlo methods, this method captures critical information about the uncertainties by considering only a relatively small number of scenarios.

As the impact of wind uncertainty is more significant than the one associated to the system demand and because of the latter is normally forecasted with enough accuracy, system demand is assumed to be perfectly forecasted and therefore it does not represent a source of uncertainty in this model. The model in [57] evaluates the wind output and the associated uncertainty. Regarding the uncertainty introduced by wind generation, the normalized wind level is assumed to follow a

Gaussian AR(2) process with half-hourly time step, which is then transformed into a non-Gaussian power output with a range from zero to the installed capacity of wind fleet.

Generation outages are assumed to follow Markov process as described in [57]. The failure of unit g may occur at each time step Δt with probability $\lambda_g \Delta t$ if the unit is committed. Unit g that is on outage is repaired with probability $\mu_g \Delta t$. We refer the reader to [57] for more details on the probability distribution of the generation outages and on the wind level distribution.

The objective of the stochastic scheduling is to minimize the expected system operation cost:

$$\sum_{n \in N} \pi(n) \left(\sum_{g \epsilon \mathcal{G}} C_g(n) + \Delta\tau(n) C_{LS} P_{LS}(n) \right) \tag{7.13}$$

subject to a load balance constraint:

$$\sum_{g \in \mathcal{G}} P_g(n) + P_W(n) - P_{WC}(n) + P_{LS}(n) = P_D(n) + P_T(n) \tag{7.14}$$

and local constraints for the thermal and storage units. We refer the reader to [57] for details on these constraints (e.g. minimum up/down time) and the equations describing generation cost functions. Since we are primarily focused on the scheduling of fast frequency response, a single bus-bar network model is used.

Given the system regulation requirements targets ($P^P(n)$ for primary, $P^S(n)$ for secondary and $P^R(n)$ for reserve), the corresponding constraints can be modelled as:

$$\sum_{g \in \mathcal{G}} R_g^P(n) + P_T^P(n) \geq P^P(n); \tag{7.15a}$$

$$\sum_{g \in \mathcal{G}} R_g^S(n) + P_T^S(n) \geq P^S(n); \tag{7.15b}$$

$$\sum_{g \in \mathcal{G}} R_g^R(n) \geq P^R(n) \tag{7.15c}$$

Note that in (7.15c) the demand side support is not considered as TCLs do not provide this service; however the impact of TCLs recovery $P^{ar}(n)$ will be taken into account in the formulation of the reserve requirement $P^R(n)$.

7.3.1 Generators' response and reserve characteristics

The primary $R_g^P(n)$ and secondary response $R_g^S(n)$ characteristics of the generating units are modelled according to the machine load level:

$$0 \leq R_g^P(n) \leq N_g^{up}(n) R_g^{P,max} \tag{7.16a}$$

$$R_g^P(n) \leq f_g^P N_g^{up}(n) \left(P_g^{max} - P_g(n)\right) \tag{7.16b}$$

$$0 \leq R_g^S(n) \leq N_g^{up}(n) R_g^{S,max} \tag{7.17a}$$

$$R_g^S(n) \leq f_g^S N_g^{up}(n) \left(P_g^{max} - P_g(n)\right) \tag{7.17b}$$

The provision of frequency response services is illustrated in Figure 50. If the unit load level is below the minimum stable generation point, clearly there is no response available as the generator is either not committed; if the unit output lies in a certain range of the unit's nominal capacity (50%-65%), the generator grants its maximum response capability. Finally, when the machine' production is approaching the maximum generation capacity, the available response decreases linearly following the slope f_g^P or f_g^S.

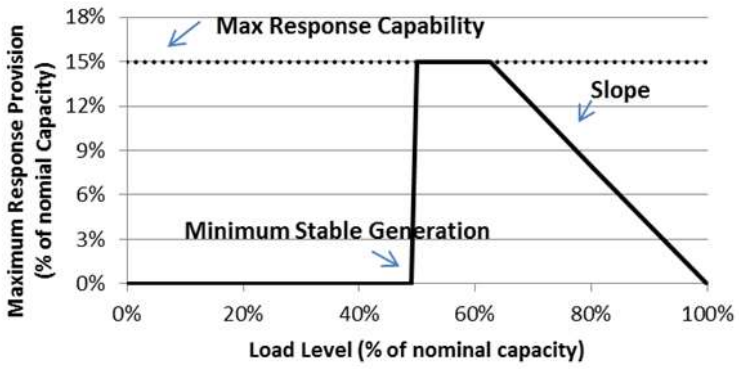


Figure 50 Example of response characteristic of conventional thermal plants.

Constraints (7.16b) and (7.17b) suggest that the same spinning headroom is allowed for primary and secondary response delivery, as current GB practice [93]. Furthermore, in this analysis the generator primary response is assumed to be a linear function of time ([94] and [95]) and thus described by a fixed slope $\frac{R_g^P}{t_p}$ until the scheduled response R_g^P is delivered. Hence, the generator power increment due to primary response provision can be modelled as:

$$\Delta P_g(t) = \frac{R_g^P}{t_p} * t \tag{7.18}$$

The contingency reserve characteristic $R_g^R(n)$ of conventional generator is modelled as the power increase at its maximum ramp rate until the predefined delivery time (t_r), the amount of reserve provided is also bounded by the spinning headroom of each generator:

$$0 \leq R_g^R(n) \leq r_g^{max} * t_r \tag{7.19a}$$

$$R_g^R(n) \leq N_g^{up}(n)\left[P_g^{max} - P_g(n) - \max\left(\frac{R_g^P(n)}{f_g^P}, \frac{R_g^S(n)}{f_g^S}\right)\right] \tag{7.19b}$$

Constraint (7.19b) requires that the spinning headroom scheduled for response should not be used for fast reserve in order to allow the recovery of response provision for the next time step.

7.3.2 Incorporating the demand side response model characteristics into the SUC

In this section, the constraints presented in Section 7.2.3, associated with the TCLs' dynamics, are incorporated into the SUC. In particular, the energy level at the end of the time period[14] corresponding to node n can be expressed as

$$S_T(n) = S_T(a(n)) \cdot e^{-\frac{\Delta t}{\hat{\tau}}} + P_T(n)\hat{\tau} \cdot \left(1 - e^{-\frac{\Delta t}{\hat{\tau}}}\right) \tag{7.20}$$

where $a(n)$ is the parent node n or rather the node that preceded node n in a generic scenario considered. The energy level and consumption level of TCLs are constrained by (7.2) and (7.5) but different bounds ($[\hat{S}_{min}^r, \hat{S}_{max}^r]$) will be used in (7.5) for regular energy levels, reflecting the need to maintain stricter temperature limits in everyday use. As in Chapter 6, the regular energy interval is 75% of this range with respect to $\hat{S}_0$. In addition, the response services provision and the consequent payback are constrained by (7.7-7.11). In Section 7.2.1 we explained the need for the TCLs average energy over 24 hours to remain close to the steady state value $\hat{S}_0$.

Let us first define $n_{\theta,\theta+i}^{(\sigma)}$ as the node at time step $t = \theta + i$, included in the scenario σ of the scenario tree, with root at time step $t = \theta$ and. Figure 51 helps to understand the notation introduced. The red arrow is pointing at the node located at time step $t = \theta + 1$, included in the

[14] The power consumed during each time step of the optimization is assumed to be constant as in Chapter 5 (see also Figure 47).

scenario $\sigma = 1$, with root at time step $t = \theta$. The associated equality constraint (7.21) follows from the demonstration of equation (5.32) in the appendix to Chapter 5 (see equation A5.8) and it is in accordance with the notation introduced in this section.

$$\frac{1}{w}\left[\varphi S\left(n_{\theta,\theta}^{(\sigma)}\right) + \sum_{i=1}^{w-1} S\left(n_{\theta,\theta+i}^{(\sigma)}\right) + \chi S\left(n_{\theta,\theta+w}^{(\sigma)}\right)\right] = \hat{S}_0 \tag{7.21}$$

where,

$$\varphi = \left[\frac{\tau}{\Delta t} - \frac{e^{-\frac{\Delta t}{\tau}}}{\left(1 - e^{-\frac{\Delta t}{\tau}}\right)}\right] \tag{7.22a}$$

$$\chi = \left[\frac{1}{\left(1 - e^{-\frac{\Delta t}{\tau}}\right)} - \frac{\tau}{\Delta t}\right] \tag{7.22b}$$

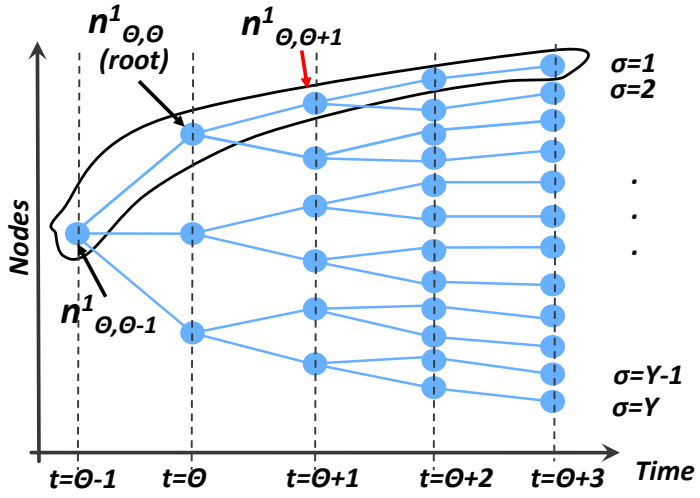


Figure 51 Node's identification: the red arrow points at a node located at time step $t = \theta + 1$ that is included in the set of nodes of the scenario $\sigma = 1$; the root of this set is at time step $t = \theta$.

It is worth pointing out that $\varphi + \chi = 1$; furthermore, in Chapter 5 and 6, because of the application of the periodic boundary conditions, it resulted that $S\left(n_{\theta,\theta}^{(\sigma)}\right) = S\left(n_{\theta,\theta+w}^{(\sigma)}\right)$ and thus the equation equivalent to (7.21) took a more compact form. However, now the whole optimization horizon is one year long, while the SUC schedules the system on a daily basis, with w=48 energy levels each time; thus in general $S\left(n_{\theta,\theta}^{(\sigma)}\right) \neq S\left(n_{\theta,\theta+w}^{(\sigma)}\right)$. At time $t = 1$, the system is scheduled taking into account constraint (7.21); the w energy levels in red in Figure 52a represent the optimal TCLs energy levels for the first 24 hours.

The application of rolling planning implies that, at time step $t = 2$, the system is rescheduled for the time window of length w that goes from $t = 2$ to $t = w + 1$; new optimal TCLs energy levels

are calculated (Figure 52b). As consequence of this, the implementation of constraint (7.21) in this new time window would not recognize the TCLs state of charge at $t = 1$ (already fixed). In general, it results that the devices would always have the possibility to postpone the energy recharge required to actually satisfy constraint (7.21). As energy and temperature are proportional quantities, TCLs will always be *on average* warmer (with cold appliances in mind).

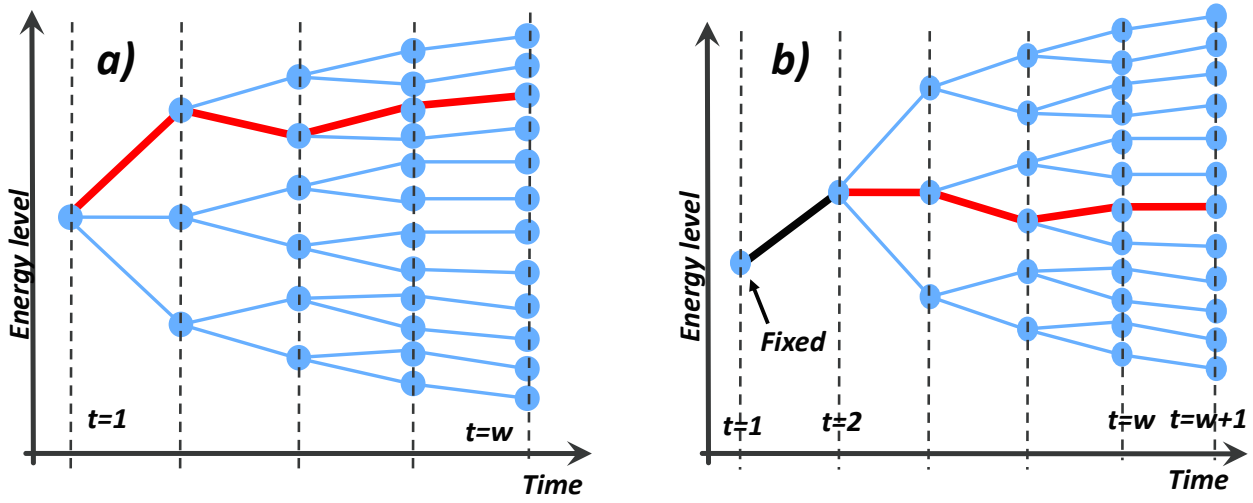


Figure 52 Schematics of a typical scenario tree in a stochastic unit commitment.

Hence, we modify constraint (7.21) so that the optimal solution at each time step is 'aware' of the energy levels already reached by the TCLs.

$$\frac{1}{w+p}\left[\varphi S\left(n_{\theta,\theta-p}^{(\sigma)}\right)+\sum_{i=1-p}^{w-1} S\left(n_{\theta,\theta+i}^{(\sigma)}\right)+\chi S\left(n_{\theta,\theta+w}^{(\sigma)}\right)\right]=\hat{S}_0 \tag{7.23}$$

In this case, at the generic time step $t = \theta$ the system is still solved over the following w time steps, but the average energy constraint takes into consideration also all p 'past' energy levels, that are not variables but fixed quantities. The impact on system cost savings and on the TCLs quality of the service of modifying the average energy constraint (7.21) with (7.23) due to the rolling planning is illustrated in Section 7.5.6.

7.4 Evaluating response and reserve system requirements

Although a precise assessment of the whole transient evolution of the frequency recovery is out of the scope of scheduling routines, in a system with a reduced level of inertia, the fast transient evolution of frequency has to be carefully considered. Hence, this section proceeds to calculate the

system requirements associated to two inertia-dependent constraints integrated in the SUC. The former calculates the minimum system inertia that limits the RoCoF below a security limit; the latter quantifies the primary response requirement to be fulfilled by generators and TCLs in order to secure the frequency nadir. In addition, we quantify the secondary response requirement for generators and TCLs based on the maximum frequency deviation at the intermediate steady state, when the primary response service is concluded. Finally we consider the reserve power to commit in order to replace the power supplied by responsive generators and to assist the TCLs energy recovery.

7.4.1 Fast frequency control and response requirements

The aim of frequency control is to contain the dynamic evolution of frequency (e.g. following a generator outage) within defined security thresholds. In Great Britain, this is specified by the GB-SQSS [20], which determines the limits on frequency deviation for secured faults. Three criteria are used to set security standards for the initial transient evolution of frequency (Figure 53):

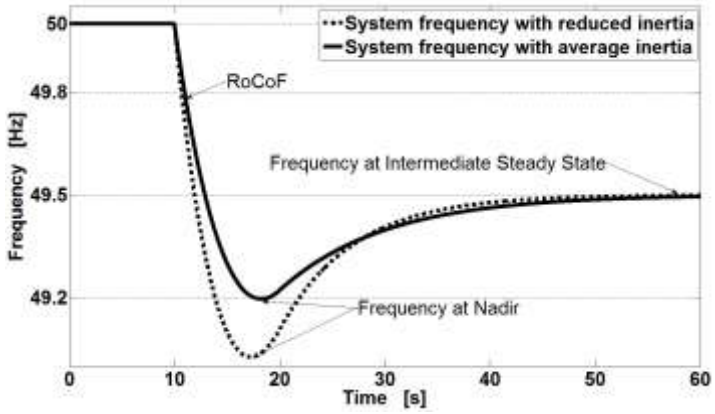


Figure 53 System frequency evolution after a contingency; the solid black evolution respects the GB security standards, while with reduced system inertia (dotted) these standards may be violated.

The RoCoF achieves the highest absolute value just after the disturbance occurs; in this narrow time window the frequency drop is only limited by the inertial response of conventional generators; currently the standard prescribes that the maximum RoCoF should not exceed 0.125Hz/s [96]. Furthermore, the generators' governor response has to limit the frequency drop to a minimum value (nadir) set to 49.2 Hz in case of the largest infeed loss [20]. The initial provision of secondary frequency response from generators enables meeting the intermediate steady state condition; in the case of GB, the frequency should stabilize above 49.5 Hz within 60s following the outage. The growing concern is that the effects of reduced system inertia (dotted curve) may

compromise the compliance with frequency regulation requirements. In particular, the RoCoF will increase and, thus, disconnections of distributed generators may be triggered by means of RoCoF-sensitive protection schemes; this action would increase the generation shortage and exacerbate the problem. Moreover, if frequency drops rapidly, conventional generators may not be fast enough to provide their primary response; the resulting frequency nadir would potentially activate the costly LFDD [91].

The time evolution of system frequency deviation can be described by a first order ODE [58]:

$$2H\frac{\partial \Delta f(t)}{\partial t} + DP_D\Delta f(t) = \Delta P_{net}(t) - \Delta P_L \tag{7.25}$$

where H [MWs^2] is the system inertia, D [%/50Hz] represents the load damping constant, P_D[MW] is the load level and ΔP_L [MW] is the generation loss; ΔP_{net} [MW] describes the additional power providing frequency response by generators and thermostatic loads. In [94] and [95] a conservative approach is adopted and damping rate is set to zero; this choice facilitates the derivation of analytical frequency response constraints but it may lead to overscheduling the amount of response requirement, thus increasing the system operational cost.

7.4.2 Rate of change of frequency

The time scale that involves the RoCoF constraint is limited to the first 1-2 seconds after the generation outage. In this context, the governor responses are still not totally triggered ($\Delta P_{net} \cong 0$) as the frequency deviation is approximately nil ($\Delta f \cong 0$). Hence, the maximum value of the rate of change of frequency ($RoCoF_{max}$) is proportional to the power shortage and inversely proportional to the system inertia; this suggests that the minimum level of system inertia $H(n)$ at each node of the scenario tree required to comply with the maximum RoCoF requirement, is obtained as follows:

$$H(n) = \frac{\sum_{g\in\mathcal{G}} H_g * P_g^{max} * N_g^{up}(n)}{f_0} \geq |\frac{\Delta P_L^{max}}{2RoCoF_{max}}| \tag{7.26}$$

where H_g is the constant of inertia [s] of each individual generator g, $P_g^{max}$is its capacity [MW], f_0 [Hz] is nominal frequency and $\Delta P_L{}^{max}$ the maximum generation expected outage. Note that wind does not provide inertial and governor response.

7.4.3 Frequency nadir

The frequency nadir is defined as the minimum value achieved by frequency during the transient period; generally, this condition is generally achieved between 2 and 10 seconds after the disturbance occurrence. The nadir depends on system inertia and governors' response; the system is assumed to be at nominal frequency (50Hz) during the pre-contingency state [57]. According to the current GB practice, the total amount of power supplied from generators providing primary response has to be delivered within t_p after the contingency [12]. Considering (7.18) and assuming that also the provision of primary response from TCLs is a linear function of time, we obtain

$$\Delta P_{net}(t) = \frac{\sum_{g\in\mathcal{G}} R_g^P + P_T}{t_p} * t = \frac{R}{t_p} * t \tag{7.27}$$

with $R = \sum_{g\in\mathcal{G}} R_g^P + P_T$, the total primary response offered by conventional generators and TCLs. The integration of equation (7.25) considering (7.27) gives the frequency deviation evolution

$$\Delta f(t) = \left(\frac{\Delta P_L}{D} + \frac{2R * H}{t_p * D^2}\right) \cdot \left(e^{-\frac{D}{2H}t} - 1\right) + \frac{R * t}{t_p * D} \tag{7.28}$$

The time t^* when the frequency reaches its nadir can be calculated by setting $\frac{\partial \Delta f(t)}{\partial t} = 0$:

$$t^* = \frac{2H}{D} \log\left(\frac{2R * H}{t_p * \Delta P_L * D + 2R * H}\right) \tag{7.29}$$

The value of frequency deviation at nadir is obtained by substituting (7.29) into (7.28). Moreover, considering the maximum generation expected outage $\Delta P_L = {\Delta P_L}^{max}$, we calculate the primary response requirement $P^p(n)$ so that at each node n the maximum frequency deviation $|\Delta f_{nadir}(n)|$ does not exceed the predefined threshold Δf_{max}.

$$|\Delta f_{nadir}(n)| = \frac{{\Delta P_L}^{max}}{D} + \frac{2P^p(n) * H(n)}{t_p * D^2} log\left(\frac{2P^p(n) * H(n)}{t_p * D * {\Delta P_L}^{max} + 2P^p(n) * H(n)}\right) \leq \Delta f_{max} \tag{7.30}$$

Rearranging equation (7.30), it gives:

$$\frac{2P^p(n) * H(n)}{t_p} \cdot log\left(\frac{2P^p(n) * H(n)}{t_p * D{\Delta P_L}^{max} + 2P^p(n) * H(n)}\right) \leq D^2 \Delta f_{max} - D{\Delta P_L}^{max} \tag{7.31}$$

Proposition:

$|\Delta f_{nadir}(n)| \leq \Delta f_{max}$ if the following mixed integer linear constraints are satisfied:

$$\begin{cases} \dfrac{\sum_{g\in\mathcal{G}} H_g * P_g^{max} * y_g(n)}{50} \geq k^* \\ -M(1-N_g^{up}(n)) \leq y_g(n) - P^p(n) \leq M(1-N_g^{up}(n)) \\ -M * N_g^{up}(n) \leq y_g(n) \leq M * N_g^{up}(n) \end{cases} \tag{7.32}$$

where *M* is a large number and $k^*(n)$ is the unique solution from

$$\frac{2k^*(n)}{t_p} log\left(\frac{2k^*(n)}{t_p * D{\Delta P_L}^{max} + 2k^*(n)}\right) = D^2\Delta f_{max} - D{\Delta P_L}^{max} \tag{7.33}$$

Proof:

The left-hand side of inequality (7.31) is a monotonically decreasing function of $P^p(n) * H(n)$ (positive quantity). Therefore, for any given value of D and ΔP_L^{max}, there exists only a unique value $k^*(n)$, such that

$$\frac{2k^*(n)}{t_p} log\left(\frac{2k^*(n)}{t_p * D{\Delta P_L}^{max} + 2k^*(n)}\right) = D^2\Delta f_{max} - D{\Delta P_L}^{max} \tag{7.34}$$

then condition $|\Delta f_{nadir}(n)| \leq \Delta f_{max}$ is therefore satisfied if

$$P^p(n) * H\,(n) \geq k^*(n) \tag{7.35}$$

The system inertia can be calculated by using

$$H(n) = \frac{\sum_{g\in\mathcal{G}} H_g * P_g^{max} * N_g^{up}(n)}{50} \tag{7.36}$$

The requirement on frequency nadir can be therefore formulated as bilinear constraint

$$\frac{(\sum_{g\in\mathcal{G}} H_g * P_g^{max} * N_g^{up}(n)) * P^p(n)}{50} \geq k^*(n) \tag{7.37}$$

By defining an additional variable $y_g(n) = P^p(n) * H(n)$ and applying standard reformulation method as in [97], condition (7.37) can be transformed into the MIL constraints (7.32).

■

The amount of power $\tilde{P}$ to be provided in order to contain $|\Delta f_{nadir}(n)| \leq \Delta f_{max}$ can be also calculated 'statically' by imposing $\frac{\partial \Delta f(t)}{\partial t} = 0$ and $\Delta f(t) = \Delta f_{max}$;

$$\tilde{P}(n) = \Delta P_L^{max} - DP_D(n)\Delta f_{max} \tag{7.38}$$

Equation (7.38) only calculates the amount of power that should be delivered by the time frequency achieves the nadir without considering the actual dynamics of the provision of primary response from generators and TCLs. Hence let us consider Figure 54; the frequency nadir is actually limited below Δf_{max} only if $\tilde{P}(n)$ is actually delivered by t^*seconds (the time when frequency achieves the nadir). Therefore, setting $P^p(n) = \tilde{P}(n)$ (see the dotted slope in Fig. 54) would not secure the frequency nadir; in fact, the primary response requirement $P^p(n)$ defines the amount of power that generators and TCLs have to be able to deliver by t_p seconds considering a power provision linear with time. (see equation 7.27)

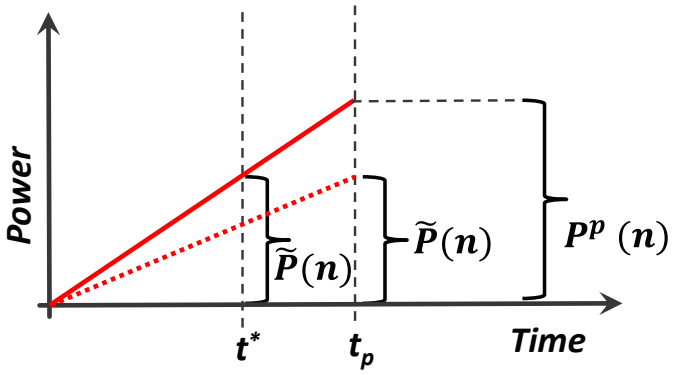


Figure 54 Schematics of the provision of primary response.

Constraint (7.35) recognizes the dynamics of frequency and primary response and, therefore, at t_p it requires $P^p(n)$ so that $\tilde{P}(n)$ is actually delivered at t^* and thus $|\Delta f_{nadir}(n)| \leq \Delta f_{max}$. The opposite situation occurs when $t^* > t_p$ and it is illustrated in Figure 55.

The primary response requirement $P^p(n)$ calculated with (7.35) would not secure the frequency nadir as generators and TCLs do not provide more response after t_p; the primary response requirement is now driven by constraint (7.38). Therefore in this case the amount of power $P^p(n)$ to be delivered by t_p is equal to $\tilde{P}(n)$, calculated with (7.38). As a consequence, if the generator outage actually occurs, frequency will reach the nadir earlier than t^*; in fact, the same amount of power $\tilde{P}(n)$ is provided faster (dotted slope). In summary, in order to guarantee that

$|\Delta f_{nadir}(n)| \leq \Delta f_{max}$, the amount of primary response provided by generators and TCLs has to be equal to

$$\sum_{g \in \mathcal{G}} R_g^P(n) + P_T^P(n) \geq P^P(n); \tag{7.39}$$

where $P^P(n)$ is the maximum value obtained from (7.35) and (7.38).

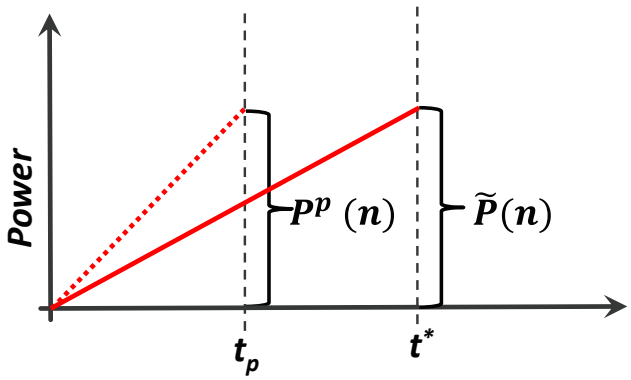


Figure 55 Provision of primary response if $t^* > t_p$.

7.4.4 Secondary response requirement

The provision of secondary response from generators permits to stabilize the frequency deviation at least at maximum intermediate steady state value Δf_{max}^{iss}; the combined action of secondary response and fast reserve brings frequency back to its nominal value. Hence, considering the maximum generation loss ΔP_L^{max}, the intermediate steady-state frequency deviation is obtained, by assuming in (7.25) that RoCoF is effectively zero (i.e. the frequency has reached a constant level):

$$|\Delta f^{iss}| = \frac{\Delta P_L^{max} - P^S(n)}{DP_D(n)} \leq \Delta f_{max}^{iss} \tag{7.40}$$

The required secondary response target, $P^S(n)$, that satisfies the intermediate steady-state frequency criterion is:

$$P^S(n) = \Delta P_L^{max} - DP_D(n)\Delta f_{max}^{iss} \tag{7.41}$$

Unlike the other two constraints, the intermediate steady-state constraint does not depend on the system inertia.

7.4.5 Reserve requirement

After the full deployment of fast reserve service, frequency is back to 50 Hz and the pre-fault generation balance is completely restored. To achieve this situation, reserve generators have to replace frequency responsive plants so that their response capability can be restored; moreover, the TCLs´ energy recovery also affects the required reserve. In fact the extra power absorbed by TCLs is supplied by reserve generators. Hence the scheduled reserve has to respect the following requirement:

$$P^R(n) \geq \Delta {P_L}^{max} + P_{\mathrm{T}}^{\mathrm{ar}}(n) \tag{7.42}$$

7.5 Case study and results

Simulations of annual system operation are performed using the GB 2030 scenario [57]. The maximum demand is 59.4 GW, the total conventional generation capacity is 70GW and the installed wind capacity is assumed to be 35GW, corresponding to a30% of wind penetration. Table 13 summarizes the characteristics of conventional plants as in [57]; wind farms do not provide inertial response and do not supply frequency services. C_{LS} is set at 30000£/MW; t_P (5 seconds) and the maximum RoCoF (0.5 Hz/s) reflect National Grid proposals for future low carbon system [12] ; in fact, current settings are t_P=10s and the maximum RoCoF (0.125 Hz/s). The load damping rate is 1 (in units of 1/50Hz) as in [58]. The maximum frequency deviation is set at 0.8 Hz and the maximum frequency deviation at the intermediate steady state, Δf_{max}^{iss}, is 0.5 Hz [12]. The duration of each time step of the SUC is Δt_1= 30 minutes; we set Δt_2= 10 minutes, and finally we select $\Delta t_3 = \Delta t_4 = \Delta t_2$. The impact of varying this setting is investigated in Section 7.5.4

Table 13 Characteristics of thermal plants

	Nuclear	***Coal***	***CCGT***	***OCGT***
Number of units	6	40	70	25
Rated power (MW)	1800	500	500	200
Minimum stable generation (MW)	1800	250	250	50
Constant of Inertia [s]	6	5	5	5
No-load cost (£/h)	0	3364	7809	8000
Marginal cost (£/MWh)	10	72	51	110
Startup cost (£)	n/a	32000	90000	0
Startup time (h)	n/a	6	4	0
Min down time (h)	n/a	4	4	0
Max primary response (MW)	0	50	50	20

	Nuclear	*Coal*	*CCGT*	*OCGT*
Primary response slope	0	0.3	0.4	0.6
Max secondary response (MW)	0	50	50	20
Secondary response slope	0	0.3	0.4	0.6
Ramp rate (MW/min)	0	15	25	35
Forced outage probability	0.02%	0.2%	0.2%	0
Mean time to repair(days)	30	3	3	n/a

55 millions of domestic fridge-freezers are considered; this number reflects the increased TCLs´ penetration forecasted for 2030 [98]. Their model parameters are listed in Table 8 (Chapter 3). Afterwards, these representative parameters are independently varied by +/-10% to generate a heterogeneous set of appliances for that class. Aggregate envelope parameters are computed using equations (5.13)-(5.14); in particular, $\hat{\mathrm{S}}_0 = 9.8\ GWh$, $\hat{\mathrm{P}}_0 = 2.8\ GW$, $\hat{\tau} = 4.51\ h$, $\hat{\mathrm{S}}_{max} = 10.56\ GW$, $\hat{\mathrm{S}}_{min} = 9.06\ GWh$, $\hat{\mathrm{P}}_{min} = 1182\ MW$ and $\hat{\mathrm{P}}_{max} = 5118\ MW$.

7.5.1 System operation cost savings with DSRM

This section explores the value of scheduling the system incorporating the DSRM proposed in this chapter. Hence, three scheduling methods are compared to the inflexible case (so-called S_ID) in which TCLs are treated as regular loads, i.e. characterized by a constant consumption and no response service provision. All of these methods share the SUC formulation in Sections 7.3-4, whereas the TCLs´ contribution is different.

1) Scheduling with constant response/no recovery (S_CRNR): the characteristics of this method are explained in Section 7.2.4; the energy and power consumption is fixed and there is no possibility to absorb extra power during the TCLs´ energy recovery phase. For this reason, the cluster of TCLs which are able to provide secondary response is limited to a fraction (1/12) of the whole devices' population.

2) *Scheduling with DSRM (S_DSRM):* this method incorporates the demand side characteristics and constraints provided in Sections 7.2 and 7.3.

4) Scheduling with flexible response ignoring the recovery (S_FRIR): similar to S_DSRM but it ignores the effect of energy recovery; generators are not scheduled to provide additional reserve ($P^R(n) \geq \Delta {P_L}^{max}$) , therefore the secondary response from TCLs results to be a cost-free service as in [43, 44].

The annual operational cost and the percentage of wind curtailment with the S_ID are 12.5 b£ and 9.7%, respectively. The performances obtained with the three methods listed above are shown in Figure 56. The cost and wind curtailment savings are referred to the S_ID scenario.

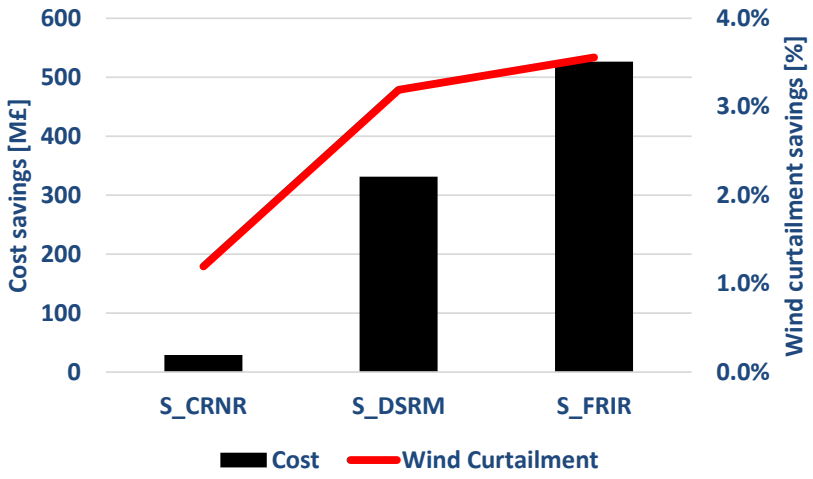


Figure 56 Annual operational cost savings (black bars) and wind curtailment savings (red line).

The cost savings provided by S_CRNR are marginal, due to inflexible energy consumption and limitations in the secondary response provision; this second aspect, in turn, is due to the absence of extra power absorption during the energy recovery phase. The S_CRNR method turns out to be not cost effective as it maintains at all times a constant energy buffer (inflexible), which is much higher compared to the actual maximum energy deployment during secondary response provision. It follows that most of the significant savings obtained with the proposed method (S_DSRM) derive from the introduction of the energy payback that enables a much larger participation of secondary response from TCLs at the expense of higher reserve requirements. Each individual TCL (55 million in total) would eventually save 6.01£ in one year as result of their flexible scheduling. The largest savings are reached by the S_FRIR method; this confirms that, ignoring the cost of TCLs, energy recovery results in over-estimating the TCLs´ value. However, this method is highly unreliable, as generators would not be able to follow the demand recovery, thus causing another frequency drop.

Let us now recall the characteristics of the DSRM discussed in Section 7.2.2. Figure 57 demonstrates the benefits of the flexible consumption/response and the inclusion of energy recovery; the graph represents the system operation for 36 hours.

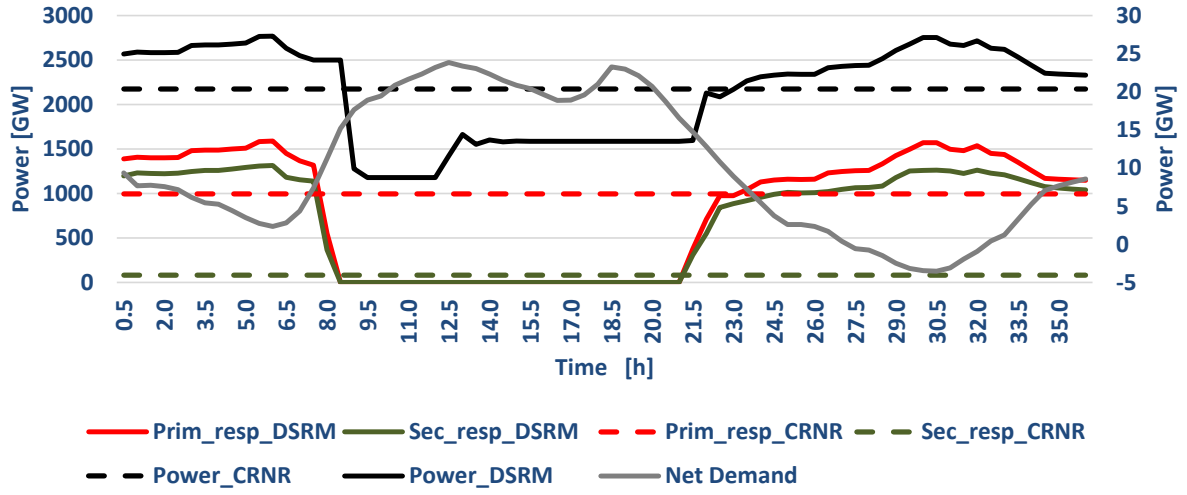


Figure 57 Actual consumption (black- solid DSRM, dashed CRNR), primary (red- solid DSRM, dashed CRNR) and secondary (green- solid DSRM, dashed CRNR) response allocation for TCLs. Net demand in grey.

The grey solid line is the net demand. The black lines illustrate the TCLs´ actual power consumption, the red lines the available primary response and the green ones the available secondary response; solid lines refer to S_DSRM, the dashed to S_CRNR. It is clear how the S_DSRM allocates more primary and secondary response when the net demand is low (i.e. high frequency requirements); note that the actual consumption is much higher than the steady state (hours 1-8.5 and 23-35). Conversely, when the net demand is high (i.e. low security requirements), the allocation of response services drops, as well as the power consumption (hours 8-21.5).

This behaviour shows the synergies between energy arbitrage and flexible response provision previously mentioned; in fact, when the net demand is low, the energy cost would be typically low and TCLs would generally absorb more energy. Moreover, the cost is held low only if TCLs also provide large amount of response (high system requirements), otherwise supplied by committing generators. The synergy is established as high consumption enables high amount of response available. On the other side, high energy costs correspond to high net demand situations; TCLs reduce the consumption facilitated by small response requirements.

In the S_CRNR scenario, the response is kept available even when it is not needed, and cannot be increased when it would be largely beneficial. Moreover, the absence of extra power absorption during the recovery permits only a marginal contribution of secondary response. It is worth pointing out that these limitations due to the energy recovery are not applied to the primary response provision. In fact, the maximum available primary response is much higher although constant. The importance of an equal availability for primary and secondary response is highlighted in the next sub-section.

7.5.2 Individual and simultaneous provision of response services

This section investigates the benefits of providing individual or simultaneous frequency services from TCLs. It emerges, from Case A (S_DSRM) in Figure 58, that it is more effective if the devices provide both services together, since there would be no value for the system if the services are individually supplied. This result reflects the comparable amplitude of the primary and secondary response requirements. Providing the first service only, it implies that many generators would still have to be committed to fulfil the secondary response requirement (being also able to supply primary control). This is in contrast with the aim of demand side response which tries to de-commit part-loaded generators to make a more efficient network operation.

The system operation shown in Figure 57 confirms this characteristic; in fact, primary and secondary response are either both committed, to a similar amount, either both not used. This result also confirms the effectiveness of facing an extra cost due to additional reserve generators (after the provision of secondary control from TCLs) rather then using the devices only for very fast frequency response and without large energy reductions.

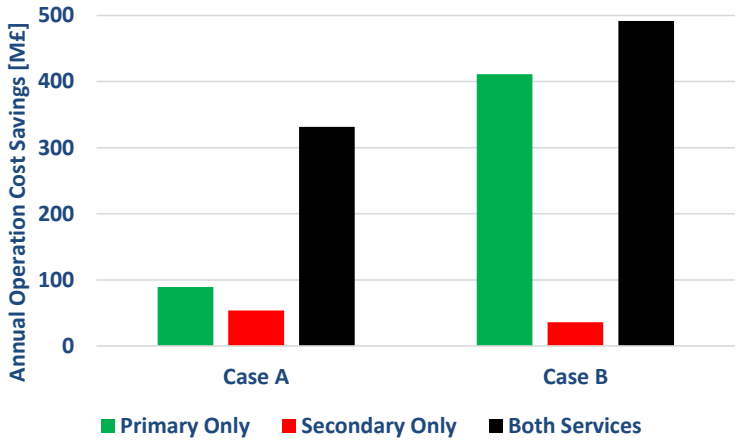


Figure 58 Response services contribution to system annual cost savings

The outcome identified may significantly differ varying generators' characteristics. NG proposals for future low carbon scenarios [12] reduced the maximum time to deliver primary response t_p from 10s (current operation) to 5s on the premise that conventional generators would be still able to deliver, by 5s, the same amount of response that is normally delivered in 10s (we adopted this assumption in the reference case (S_DSRM)). However, if generators can actually deliver by 5s only half of the response capacity due to technical limits (Case B, Figure 58), the cost benefits associated to the provision of primary response only from TCLs will not significantly differ to those achieved for

a multi-services supply. The primary response requirement is now much higher than the one of secondary response; if TCLs provide only primary control, the generators engaged to supply secondary response will be still committed to help the fulfilment of the large primary response requirement. Note that Case B considers less flexible generators thus the value for TCLs support is higher compared to the one obtained in Case A. However, the annual operational cost related to Case B (12.54 b£) is still higher than the equivalent cost for Case A (12.21 b£).

7.5.3 Sensitivity of the operational cost savings to the primary response delivery time

We now proceed to study the sensitivity of the annual operational cost savings to the primary response delivery time t_p. In particular, the settings are varied between 5s (reference case in this Chapter) up to 10s (National Grid current operation). The maximum capacity of the demand side response support at t_p does not change whereas, regarding the conventional plants, we assume that the slope $\frac{{R_g}^p}{t_p}$ of the linear generator response is constant and the maximum response capacity increases accordingly. This way we investigate the value for demand side actors to deliver fast frequency support rapidly. The results in Figure 60 show a reduction of the annual cost savings if the primary response delivery time t_p increases up to 10s.

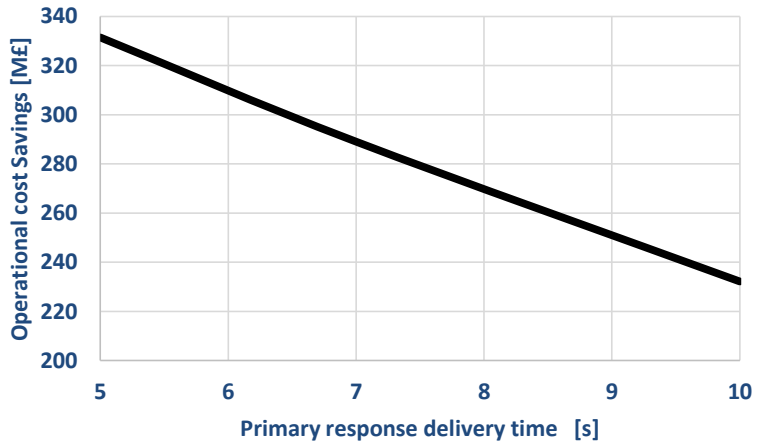


Figure 59 Impact of varying the primary response delivery time on annual system operation cost.

These results reflect the outcomes obtained in [57] and in a National Grid technical report [12]; the amount of primary response required drastically increases with delayed response provisions. It follows that the higher the delivery time setting, the larger the generators' headroom for primary response required for providing the same amount of power. As in such a network with

reduced inertia the frequency nadir would be reached very quickly (often before 5s), it results that a high delivery time setting leads to a more inefficient generators' operation (more part-loaded).

7.5.4 Sensitivity of the operational cost savings to the TCLs´ recovery pattern

The relation between the additional reserve required due to TCLs´ recovery and the secondary response from the devices is regulated by the function $L_3 = f(\Delta t_3)$, as explained in Section 7.2.3. For a given time interval Δt_1 and for a given secondary response commitment Δt_2, this function only depends on Δt_3. The sensitivity of the operational cost savings to the energy recovery pattern is shown in Figure 60 where the reference case, Case A (Δt_3=10mins), is compared with Case B (Δt_3=5mins) and Case C (Δt_3=15mins).

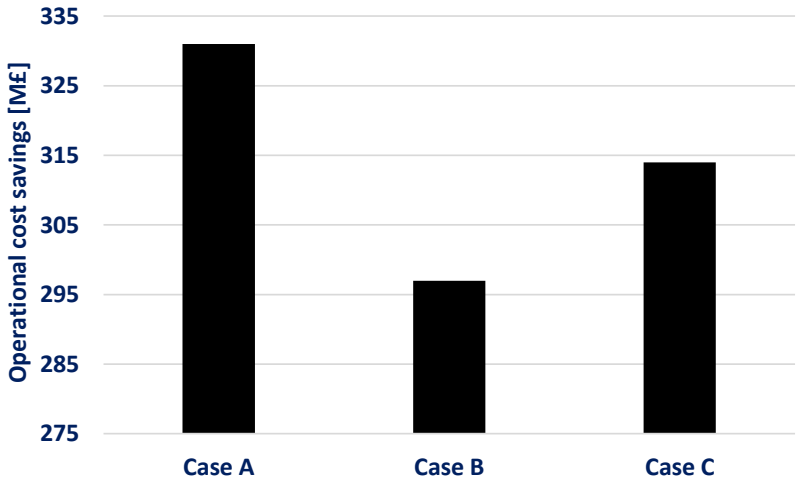


Figure 60 Sensitivity of the operational cost savings to the TCLs energy recovery pattern

The outcomes show that the best solution is the result of a trade-off between the amount of the additional reserve and the speed in the provision of this service. The average reference setting Δt_3=10mins is the most cost effective solution. Reducing Δt_3 decreases the amplitude of P_i^{ar}, but it requires a faster provision of the reserve from the generators; in fact, these units will face technical limitations (ramping constraints); the system will have to schedule more available units, thus increasing the operational cost. Conversely, with a large Δt_3, the power limitations on reserve machines due to the ramping constraints decrease. However, the amount of reserve to be provided does increase and moreover the TCLs´ secondary response actually committed now could be sometimes limited by the upper bound of (7.5); this issues cut down the advantages of this setting.

7.5.5 Impact of higher TCLs penetration

In this study we quantify the impact that a higher penetration of smart thermostatically controlled loads has on the annual system operational cost savings. For simplicity we simulate the effect of a larger penetration of TCLs just doubling the number of thermostatic units considered in the reference case. Although this simple assumption does not recognize the capabilities of different appliances (e.g. commercial refrigeration units), the results obtained in Figure 61 still suggest that a more spread demand side support will lead to significant further system cost savings.

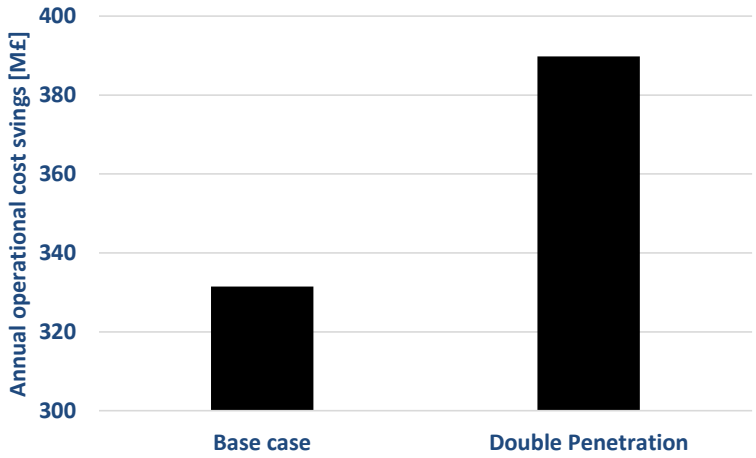


Figure 61 Annual operational cost savings due to a double TCLs penetration

With a double TCLs´ penetration, system operational cost savings increase by around 18%.

7.5.6 Discussion on the average energy constraint

This section extends the discussion in Sec. 7.3.2 regarding the need for TCLs´ average energy to remain close to its steady state $\hat{S}_0$. Three scheduling methods (all with rolling planning) are considered with different implementations of the constraint on the average state of charge (SOC). Case A implements constraint (7.21) while Case B (reference case) implements (7.23). The last option, Case C, guarantees that the average SOC of each real day of the year equals $\hat{S}_0$. In this case, the average constraint (7.23) is still applied in conjunction with (7.21) that is now implemented only to the physical 24 hours of each day; hence, Figure 62 shows that the highest cost savings, represented by the black bars, is achieved by Case A. However, this method makes an unfair use of the TCLs´ energy storage; in fact the annual average SOC is around by 4.5% lower than $\hat{S}_0$; the mean error is represented in green.

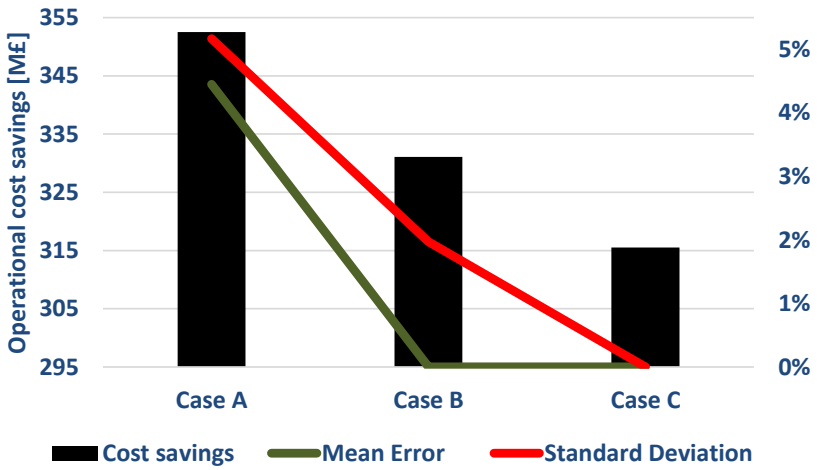


Figure 62 Comparison of the scheduling methods based on cost savings (black bars), mean error between the annual average SOC and $\hat{S}_0$ (green line) and standard deviation of the pdf of the average daily SOC (red line).

The mean error is nil for Case B and C therefore the average SOC is maintained in the long term. However, although with Case B the daily average SOC is not strictly guaranteed over each real day, the standard deviation of the pdf of the average SOC (red line) for each actual day of the year is low (1.9%). In fact, Case B permits to operate the refrigerators slightly warmer over one day, if response requirements are particularly low, and slightly colder in any other day characterized by higher system requirements. The augmented flexibility of Case B enables larger cost savings compared to stricter Case C. Note that the energy level and the temperature are in proportional relation.

7.6 Conclusions

This chapter introduces a novel DSRM for a heterogeneous population of thermostatically controlled loads. The DSRM, integrated in an advanced SUC, optimizes (a) the energy consumption of the appliances and (b) allocates the optimal amount of response services to be provided in case of generator failure. The optimal solution minimizes the annual system operational cost. The control strategy discussed in Chapter 4, with the modifications introduced in Chapter 5, ensures the feasibility of the optimal profile by means of energy and power constraints.

In Section 7.2 we identify three key ingredients that characterize the methodology developed; the first one regards the flexible TCLs´ energy consumption; in fact the appliances are able to adjust the actual energy level in order to deliver more response services during those periods with high system response requirements (i.e. low net demand). If the reverse case happens, an

overall reduction of energy/power consumption is allowed. This behaviour results to be in synergy with the possibility for TCLs to make energy arbitrage, this way increasing the consumption in presence of low energy cost (net demand low) and reducing the absorption during those period with high energy costs (high net demand).

The second characteristic deals with the TCLs´ energy recovery pattern; the proposed DSRM allows for a full controllability of this phase; during the payback period, TCLs can absorb an extra amount of power to make the energy recovery quicker, letting TCLs able to provide significant response support at the next time step. Hence the additional power is supplied by reserve generators. Therefore, the demand side value is calculated considering the cost of this increment.

The last feature deals with the reliability of the response support from TCLs; the DSRM is constructed in such a way that the devices can always guarantee the deliverability of the scheduled response regardless the actual occurrence of generators' failures at previous time step. In fact, if response provision is actually called up due to a generation outage, the energy level reached by the TCLs´ consequent payback period is the same that would have been reached following the prescheduled "regular operation" profile. This represents an important contribution as it makes the TCLs´ response provision as reliable as the generators' ability to provide frequency control.

The mathematical formulation of the DSRM and the associated constraint have been illustrated in Section 7.2.3; in this section we demonstrated that the amount of extra power required during the recovery period is only a function of the amount of secondary response allocated. Moreover, in Section 7.2.4 we focused on an alternative framework for demand side response with thermostatic loads. Under this approach, the energy consumption is inflexible and the energy recover after providing response services has to be performed without resorting to extra power consumption. These peculiar features do limit the intrinsic flexibility of TCLs, thus reducing the system cost savings obtained by integrating such a scheme for demand side response. Sections 7.3 and 7.4 are dedicated to the formulation of the SUC problem and the modifications made in order to incorporate the demand side contribution. In fact the inertia-dependent constraints permit to secure the fast transient frequency drop below security thresholds. In a future system scenario, characterized by a reduced level of inertia, the assessment of primary response requirements in a static fashion will be no longer sufficient.

Section 7.5 introduced the case study; the proposed stochastic scheduling with the DSRM is compared to the solution discussed in Section 7.2.4 in order to show the effectiveness of our

method. The comparison is based on the annual system operational cost savings and wind curtailment savings. In the following two sub-sections we performed sensitivity analysis varying the maximum response capacity of conventional generators and the settings of the primary response delivery time. From the first study it emerges the need for a simultaneous and comparable provision of primary and secondary response in order to impact the system operational cost. The second study confirms the benefits associated to the National Grid proposal to reduce the primary response delivery time from the current setting of 10s to 5s. A further study explores the value of varying the TCLs´ recovery pattern; the relation between the additional reserve required and the secondary response from TCLs is given by a function, L_3, that depends only on Δt_3 if Δt_1 and Δt_2 are fixed. Therefore, in this study we verify that the optimal setting for Δt_3 would be an intermediate value($\Delta t_3{}^* = 10$min if $\Delta t_3 = \{5,10,15\}$; small values are penalised by strict generators' rump rates constraints while, for large values, the TCLs´ consumption is limited by the maximum TCLs´ power threshold and the amount of extra reserve required does increase.

Finally in Section 7.5.6 we discussed the impact that the use of rolling planning in the scheduling routine has on the TCLs´ quality of the service. To maintain a good level of quality of the primary function of the TCLs it is important that the average energy/temperature does not deviate too much from the steady state value. The simple application of the average constraint (7.21) is not doable as the average state of charge (temperature) is quite lower (higher) than the steady state value; both the other two methods are valid in the sense that the average state of charge is maintained over the whole year. The difference between these two methods lies in the strict respect of a constraint on the average energy during the actual physical day. In the less flexible case the average energy over each day of the year is equal to the steady state energy; in the other case, there are some days during which the average energy is slightly higher and some others during which the average energy is slightly lower than the steady state value. This extra flexibility translates into higher cost savings; here we limit to point out the issue as a final assessment, also based on the quality of the services requirements, is beyond of the scope of this thesis.

7.6.1 Extensions

The work presented in this chapter has room for improvements. The first one regards the SUC; the model developed does not explicitly consider different ramp rates of individual generators in the delivery of scheduled primary response. The introduction of inertia-dependent constraints

represents a basis for incentivizing fast-generators to provide a more rapid frequency response delivery. Of course, this issue does affect the demand side response; in fact, TCLs could actually conclude the full power reduction within very short time period, as they are not limited by the mechanical generators' constraints (e.g. valve openings). Moreover, it will be important to make use of the total demand side contribution; significant support may be introduced by other TCL appliances from different sectors (heating systems).

Finally, ongoing research is focused on making the energy recovery pattern flexible; the additional power $P_{\mathrm{T}}^{\mathrm{ar}}(n)$ is related to the amount of secondary response allocated by means of function L_3. Once the duration of the time step Δt_1 (e.g. 30 minutes) and the commitment of secondary response Δt_2 are fixed, the function L_3 only depends on Δt_3 (see Section APP7.3 in the appendix to this chapter). Moreover this function L_3 is approximately linear with Δt_3. With few modifications to the SUC framework it would be possible make the duration $\Delta t_3(n)$ of the recovery period a decision variable. At each time step, the provision of extra reserve power from generators will be optimized according to the available units; for instance, fixing the secondary response provision, choosing a higher $\Delta t_3(n)$ would permit to not call up another generator even thoug the net amount of power to be provided is higher.

Appendix

APP7.1 Derivation of equations (7.9)-(7.11)

We now proceed to validate equations (7.9)-(7.11) and derive the constant coefficients introduced in these equations. Let us first define

$$\gamma_i = e^{-\frac{\Delta t_i}{\hat{\tau}}} \qquad \text{(A7.1a)} \qquad\qquad \beta_i = \hat{\tau} \cdot (1 - \gamma_i) \qquad \text{(A7.1b)}$$

Considering (7.6b), we calculate S_i^r as the energy at the end the second sub-interval in Figure 47.

$$S_i^r = \gamma_3 S_i^{\Delta} + \hat{\tau}\Delta t_3 \rho_i' + \beta_2 q_i' - \beta_2 \hat{\tau} \rho_i' \qquad \text{(A7.2)}$$

By inserting the following equations (A7.3)

$$\rho_i' = \frac{P_i^r - (P_i - P_i^s)}{\Delta t_3} \qquad \text{(A7.3a)} \qquad q_i' = P_i - P_i^s \qquad \text{(A7.3b)}$$

in (A7.2) and considering (7.9) we obtain

$$S_i^r = R_1 S_{i-1} + R_2 P_i - R_2 P_i^s + R_3 P_i^r \qquad \text{(A7.4)}$$

with,

$$R_1 = \gamma_2\gamma_3; \qquad R_2 = \gamma_3\beta_2 - \hat{\tau} + \beta_3 - \frac{\beta_3\hat{\tau}}{\Delta t_3}; \qquad R_3 = \hat{\tau} - \frac{\beta_3\hat{\tau}}{\Delta t_3} \qquad \text{(A7.5)}$$

Hence we now derive again S_i^r (A7.9) as the energy at the beginning the third sub-interval in Figure 47.

$$S_i^r = \frac{1}{\gamma_4}(S_i - \hat{\tau}\Delta t_4 \rho_i'' - \beta_4 q_i'' + \beta_4\hat{\tau}\rho_i'') \qquad \text{(A7.6)}$$

The energy S_i at the end of the interval i is obtained as follows;

$$S_i = \gamma_1 S_{i-1} + \beta_1 P_i \qquad \text{(A7.7)}$$

In this sub-interval we find that

$$\rho_i'' = \frac{P_i - P_i^r}{\Delta t_4} \qquad \text{(A7.8a)} \qquad q_i'' = P_i^r \qquad \text{(A7.8b)}$$

Therefore the energy S_i^r is

$$S_i^r = G_1 S_{i-1} + G_2 P_i + G_3 P_i^r \qquad \text{(A7.9)}$$

with,

$$G_1 = \frac{\gamma_1}{\gamma_4}; \qquad G_2 = \frac{\beta_1 - \hat{\tau}}{\gamma_4} + \frac{\beta_4\hat{\tau}}{\gamma_4\Delta t_4}; \qquad G_3 = \frac{\hat{\tau} - \beta_4}{\gamma_4} - \frac{\beta_4\hat{\tau}}{\gamma_4\Delta t_4}; \qquad \text{(A7.10)}$$

Imposing that equation (A7.4) equals (A7.9) it is possible to derive the expression of P_i^r in equation (7.10), considering,

$$L_1 = \frac{G_1 - R_1}{R_3 - G_3}; \qquad L_2 = \frac{G_2 - R_2}{R_3 - G_3}; \qquad L_3 = \frac{R_2}{R_3 - G_3} \qquad \text{(A7.11)}$$

Finally, substituting (7.10) in (A7.9) the formulation of equation (7.11) is obtained, considering

$$H_1 = G_1 + G_3 L_1; \qquad H_2 = G_2 + G_3 L_2 ; \qquad H_3 = G_3 L_3 \tag{A7.12}$$

APP7.2 Further consideration on coefficients L_1 and L_2

Proposition: the value of the coefficient L_1 is 0 and the value of coefficient L_2 is 1

Proof:

Let us recall equation (A7.11); substituting in this equation the expressions for R_1 , R_3 (A7.5), and G_1 , G_3 (A7.10), we obtain

$$L_1 = \frac{G_1 - R_1}{R_3 - G_3} = \frac{\frac{\gamma_1}{\gamma_4} - \gamma_2\gamma_3}{R_3 - G_3} = \frac{\frac{e^{-\frac{\Delta t_1}{\hat{\tau}}}}{e^{-\frac{\Delta t_4}{\hat{\tau}}}} - e^{-\frac{\Delta t_2}{\hat{\tau}}} e^{-\frac{\Delta t_3}{\hat{\tau}}}}{R_3 - G_3} = \frac{e^{-\left(\frac{\Delta t_1 - \Delta t_4}{\hat{\tau}}\right)} - e^{-\left(\frac{\Delta t_2 + \Delta t_3}{\hat{\tau}}\right)}}{R_3 - G_3} \tag{A7.13}$$

Considering Figure 47 it results that,

$$\Delta t_1 - \Delta t_4 = \Delta t_2 + \Delta t_3 \tag{A7.14}$$

therefore,

$$e^{-\left(\frac{\Delta t_1 - \Delta t_4}{\hat{\tau}}\right)} - e^{-\left(\frac{\Delta t_2 + \Delta t_3}{\hat{\tau}}\right)} = 0 \qquad \Longrightarrow \qquad L_1 = \frac{e^{-\left(\frac{\Delta t_1 - \Delta t_4}{\hat{\tau}}\right)} - e^{-\left(\frac{\Delta t_2 + \Delta t_3}{\hat{\tau}}\right)}}{R_3 - G_3} = 0 \tag{A7.14}$$

Considering again equations (A7.11), (A7.5) and (A7.10), we have

$$L_2 = \frac{G_2 - R_2}{R_3 - G_3} = \frac{\frac{\beta_1 - \hat{\tau}}{\gamma_4} + \frac{\beta_4 \hat{\tau}}{\gamma_4 \Delta t_4} - \gamma_3\beta_2 + \hat{\tau} - \beta_3 + \frac{\beta_3 \hat{\tau}}{\Delta t_3}}{\hat{\tau} - \frac{\beta_3 \hat{\tau}}{\Delta t_3} - \frac{\hat{\tau} - \beta_4}{\gamma_4} + \frac{\beta_4 \hat{\tau}}{\gamma_4 \Delta t_4}} \tag{A7.15}$$

Hence $L_2 = 1 \iff G_2 - R_2 = R_3 - G_3$. It follows that,

$$\frac{\beta_1 - \hat{\tau}}{\gamma_4} + \frac{\beta_4 \hat{\tau}}{\gamma_4 \Delta t_4} - \gamma_3\beta_2 + \hat{\tau} - \beta_3 + \frac{\beta_3 \hat{\tau}}{\Delta t_3} = \hat{\tau} - \frac{\beta_3 \hat{\tau}}{\Delta t_3} - \frac{\hat{\tau} - \beta_4}{\gamma_4} + \frac{\beta_4 \hat{\tau}}{\gamma_4 \Delta t_4} \tag{A7.16}$$

Equation (A7.16) can be easily simplified and it becomes

$$\frac{\beta_1}{\gamma_4} - \gamma_3\beta_2 - \beta_3 = \frac{\beta_4}{\gamma_4} \tag{A7.17}$$

We now reformulate (A7.17) considering (A7.1); thus

$$\frac{\hat{\tau}\left(1-e^{-\frac{\Delta t_1}{\hat{\tau}}}\right)}{e^{-\frac{\Delta t_4}{\hat{\tau}}}}-\hat{\tau}e^{-\frac{\Delta t_3}{\hat{\tau}}}\left(1-e^{-\frac{\Delta t_2}{\hat{\tau}}}\right)-\hat{\tau}\left(1-e^{-\frac{\Delta t_3}{\hat{\tau}}}\right)=\frac{\hat{\tau}\left(1-e^{-\frac{\Delta t_4}{\hat{\tau}}}\right)}{e^{-\frac{\Delta t_4}{\hat{\tau}}}} \tag{A7.18}$$

With few simplifications equation (A7.18) becomes

$$e^{\frac{\Delta t_4}{\hat{\tau}}}-e^{-\left(\frac{\Delta t_1-\Delta t_4}{\hat{\tau}}\right)}-e^{-\frac{\Delta t_3}{\hat{\tau}}}+e^{-\left(\frac{\Delta t_2+\Delta t_3}{\hat{\tau}}\right)}-1+e^{-\frac{\Delta t_3}{\hat{\tau}}}=e^{\frac{\Delta t_4}{\hat{\tau}}}-1 \tag{A7.19}$$

Recalling (A7.14) it is easy to verify that equation (A7.19) turns into the identity $0=0$.

■

APP7.3 Further consideration on function L_3

Let us substitute equations (A7.10) and (A7.5) in (A7.11).

$$L_3=\frac{R_2}{R_3-G_3}=\frac{\gamma_3\beta_2-\hat{\tau}+\beta_3-\frac{\beta_3\hat{\tau}}{\Delta t_3}}{\hat{\tau}-\frac{\beta_3\hat{\tau}}{\Delta t_3}-\frac{\hat{\tau}-\beta_4}{\gamma_4}+\frac{\beta_4\hat{\tau}}{\gamma_4\Delta t_4}} \tag{A7.20}$$

Considering (A7.1) we obtain

$$L_3=\frac{\hat{\tau}e^{-\frac{\Delta t_3}{\hat{\tau}}}\left(1-e^{-\frac{\Delta t_2}{\hat{\tau}}}\right)-\hat{\tau}+\hat{\tau}\left(1-e^{-\frac{\Delta t_2}{\hat{\tau}}}\right)\left(1+\frac{\hat{\tau}}{\Delta t_3}\right)}{\hat{\tau}\left[1-\frac{\hat{\tau}\left(1-e^{-\frac{\Delta t_3}{\hat{\tau}}}\right)}{\Delta t_3}\right]+\frac{\hat{\tau}\left(1-e^{-\frac{\Delta t_4}{\hat{\tau}}}\right)-\hat{\tau}}{e^{-\frac{\Delta t_4}{\hat{\tau}}}}+\frac{\hat{\tau}^2\left(1-e^{-\frac{\Delta t_4}{\hat{\tau}}}\right)}{\Delta t_4 e^{-\frac{\Delta t_4}{\hat{\tau}}}}}=$$

$$=\frac{\frac{\hat{\tau}}{\Delta t_3}-e^{-\frac{(\Delta t_2+\Delta t_3)}{\hat{\tau}}}-\frac{e^{-\frac{\Delta t_3}{\hat{\tau}}}}{\Delta t_3}}{\frac{\Delta t_3-\hat{\tau}\left(1-e^{-\frac{\Delta t_3}{\hat{\tau}}}\right)}{\Delta t_3}-1+\frac{\hat{\tau}\left(1-e^{-\frac{\Delta t_4}{\hat{\tau}}}\right)}{e^{-\frac{\Delta t_4}{\hat{\tau}}}\Delta t_4}} \tag{A7.21}$$

We use (A7.14) to eliminate Δt_4; thus

$$L_3 = \frac{\frac{\hat{\tau}}{\Delta t_3} - e^{-\frac{(\Delta t_2 + \Delta t_3)}{\hat{\tau}}} - \frac{e^{-\frac{\Delta t_3}{\hat{\tau}}}}{\Delta t_3}}{\frac{\hat{\tau}\left(e^{-\frac{(\Delta t_2 + \Delta t_3)}{\hat{\tau}}} - 1\right)}{\Delta t_1 - (\Delta t_2 + \Delta t_3)} - \frac{\hat{\tau}}{\Delta t_3}\left(1 - e^{-\frac{\Delta t_3}{\hat{\tau}}}\right)} \tag{A7.22}$$

Note that Δt_1 and Δt_2 are constant values and the only variable is Δt_3. Let us now plot this function using the default parameters introduced in Section 7.5.

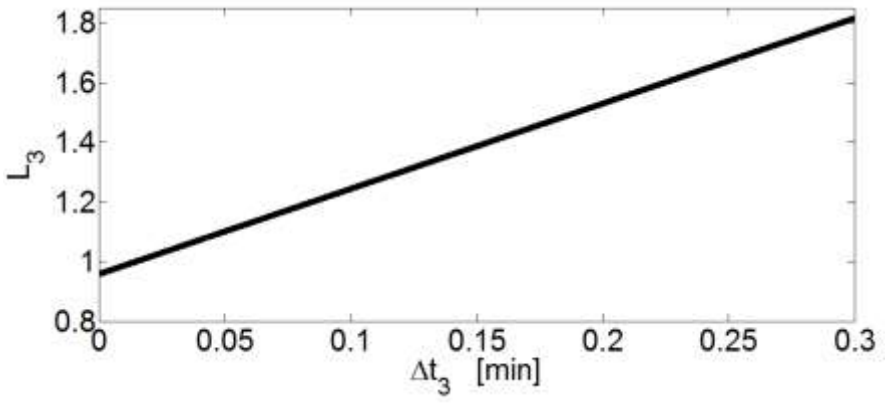


Figure 63 Function L_3 against the time interval Δt_3.

Note $\Delta t_3 = 0$ means that, from the reduced power consumption due to the deliver of secondary response, TCLs sharply reach the peak of the extra recovery consumption. This figure shows the quasi linear-relation between L_3 and Δt_3. This relation is confirmed by taking the expansion in Taylor series:

$$L_3 = 0.95766 + 2.85603\Delta t_3 + 0.00262{\Delta t_3}^2 + 0.00004{\Delta t_3}^3 + \mathcal{O}\left({\Delta t_3}^4\right) \tag{A7.23}$$

The coefficients after the linear term are in practice negligible.

Chapter 8: Conclusions and Further Work

8.1 Conclusions

The work presented in this thesis quantifies the impact that the use of flexible thermostatically controlled loads has on the operation and security of future low-carbon power systems, which are characterised by the need for significant availability of various ancillary services. The work carried out has focused on the ability and benefits of providing frequency response services and various system services (e.g. energy arbitrage and transmission congestion relief) with TCLs, an interesting category within the larger set of demand side response. The results obtained in this thesis strongly confirm the huge potential benefits for the system operation and security if a large cluster of controllable TCLs is integrated within the system. As the research in this field is substantially growing, it has to be supported by a much clearer regulation framework.

This thesis made a contribution in all the three key-aspects of the demand side response with TCLs: the modelling, the control and the implementation of TCLs models in advanced network scheduling routines. The methodology followed to achieve these results has been based on the prior achievement of two technical steps; the first deals with the derivation of dynamic thermal models to describe the temperature dynamics of various thermostatic loads (Chapter 3) in the refrigeration sector (domestic and commercial). Eight different classes of cold appliances have been modelled with high level of accuracy with regard to the scope of this work. The second was the design of an accurate control strategy and an aggregate TCL population description to govern the overall power consumption of the TCLs' cluster (Chapters 4-5). It results that a large population of heterogeneous TCLs can be accurately controlled over different time scale and described by means of a single battery-like model; hence, we have proved that arbitrary profiles (within limits) can be actually

tracked by single TCLs individually. Moreover, we have provided an aggregate leaky storage model that precisely describes the thermal energy dynamics of the TCLs but it also facilitates a straightforward embedding in any optimization tool. Chapters 6 and 7 build on these results to establish the system-level impact of flexible TCLs on the generators' dispatch (Chapter 6) and commitment (Chapter 7).

It has been demonstrated that the benefits of integrating this promising technology in the network operation result in significant system operational cost savings on an annual basis; against 55 million of TCLs (approximately 2.8 GW average power), the annual system operation cost reduces up to 331 million of pounds are saved in one year. This significant value comes from the simultaneous provision of frequency response services and the parallel ability to optimize the TCL energy consumption in a flexible way (energy arbitrage). Moreover, it has been shown that the flexible control of TCLs enables the absorption of more wind generation in the system operation; this action positively contributes to the reduction of greenhouse gases emissions, in compliance with ambitious national and international environmental targets. Concerning this aspect, the level of wind curtailment drops from more than 20% to 9.7%. Further outcomes from this thesis reveal that the energy payback period cannot be neglected in any realistic study but also it is not a negative issue for the system operation. In effect, it represents as sort of compromise to shift large shares of conventional generators' support to much longer time scales and maintain high level of reliability in the provision of fast response services from TCLs at each scheduling period. We now move from a high level perspective to summarize the work carried out and the associated conclusions obtained each individual chapter. It is worth pointing out that detailed conclusions are offered at the end of each chapter.

Chapter 2 dealt with the potential support from TCLs to the inertial response of a low carbon network characterized by a reduced amount of inertia. It has been demonstrated how a rapid and distributed load shedding of not-time critical appliances can drastically improve the inertial response of the system. We made it possible to associate a value of inertia to TCLs even though the specific definition of constant of inertia cannot be applied to these static devices. In particular, we

- proposed the concept of "effective dynamic inertia" as a concept to extend the inertia of conventional generators in presence of demand side response. This concept is general and it is not limited to thermostatically controlled loads,

- proposed the concept of "average inertia" more useful for TSOs to calculate system level security requirements,
- demonstrated that a population of thermostatic loads effectively increments the inertia of network, thus facilitating the integration of wind generation,
- compared the inertial support from TCLs with the inertia of a system characterized by only synchronous generators and we demonstrated the benefits of the first case.

Chapter 3 focused on the modelling of various thermostatic loads within the refrigeration sector (domestic and commercial) and investigated initial controllers for TCLs designed for frequency control purposes. As a consequence of the smart action of several millions of TCLs the frequency dynamics after a large generator failure remain within security thresholds even in presence of large shares of system energy supplied by wind farms (no ancillary services provided by this technology). In detail, we

- derived detailed 4^{th} order dynamic models to describe the temperature evolution of a domestic combined fridge-freezer; second order models for domestic appliances with only one compartment i.e. fridge or freezer were obtained by downgrading the general 4^{th} order model. These models are able to separate the temperature dynamics of the temperature of the internal wall of one compartment (controlled temperature) from those related to the content (food),
- derived first order models for commercial refrigerators such as bottle cooler, commercial fridges and freezers and multidecks; effective and appropriate models for these TCLs could not be found in public,
- analysed the properties of the collective response of TCLs with respect to the provision of frequency response; in particular we clearly distinguished the issue of the synchronization that results from the natural temperature recovery. This analysis defined benefits and the drawbacks associated to initial controller's designs (deterministic controller and stochastic controller),
- designed an hybrid controller that respected three system level performance criteria; in fact, it offers increased short-term power reduction as the temperature deadband that controls the TCL switching is also sensitive to the rate of change of frequency;

moreover, it largely suppresses the issue of the synchronization thanks to an effective disengagement strategy. The onset of the payback period is delayed.

In Chapter 4 we have illustrated the framework and the main outcomes of a novel class of controllers for governing the aggregate power consumption of TCLs. The control strategy developed permits for the first time to accurate control in a decentralised and non-disruptive way the temperature and power dynamics of a large population of TCLs. In particular, we

- proved that the proposed controller is able to simultaneously address four criteria that define the optimal use of demand response capability; other methods in the literature could only accomplish one or more of those but not all at the same time,
- reverted the approach of the control problem as we make use of a collective property (the heating rate) as control parameter; stochastic switching and threshold temperatures are derived only in a second phase,
- applied the general framework of the analysis to the case of TCLs described by linear first ODEs; this step provided two general results: first, it is straightforward to compute the evolution of the ensemble average temperature from the aggregate power consumption - regardless of the specifics of the controller. Second, a so-called 'payback' period of increased power consumption is not strictly required after a power reduction; it may be used to speed up the energy recovery though,
- introduced a specific controller where the control parameter (the heating rate) is function of the temperature; for this case, a comparison with the generic limitations for linear thermal models shows that the proposed structure is optimal with respect of the energy provision for load reduction,
- quantified the impact of two different implementations of the proposed controller on the system frequency control. A frequency linear controller and a pre-programmed controller provide promising results towards the integration of RES in a future power system.

Chapter 5 introduces the leaky storage model to describe a population of TCLs; this battery-like model has the benefit to allow a straightforward embedding in any optimization tools. Thus, the studies developed address the simultaneous and optimal allocation of various response services and

in conjunction with the optimal daily energy consumption that maximise the revenues for a portfolio of TCLs. The results show the promising economic benefits for demand side response with TCLs. In particular, in this chapter we

- demonstrated how a large heterogeneous population of thermostatically controlled loads can be described by a *leaky storage unit*, an insightful representation from the power system point of view. The accessible storage levels $S(t)$ and power consumption levels $P(t)$ of this unit are limited by the appliances' thermal models, their temperature constraints and the control strategy adopted,
- showed that leaky storage model is exact for large populations of TCLs ($N \to \infty$) with linear thermal models and governed by the control methodology discussed in Chapter 4,
- extended of the advanced controller proposed in Chapter 4 by matching the power-reduction only design with its power-increasing counterpart. In this way we achieved full use of the available temperature interval,
- embedded the leaky storage model in a revenue-maximising scheduling model. This enables an optimal allocation of energy and ancillary services that takes into account the physical properties and the constraints associated with the quality of the service of the appliances and possible conflicts or synergies between the services considered,
- showed that, for the classes of TCLs considered, very different service allocations result from differences in models' parameters. The allocations largely reflect the approximate analysis carried out by means of dimensionless parameters introduced in Section 5.3.

In Chapter 6 we changed the perspective of the analysis and we focused on the system level impact of a population of TCLs. The obtained results show a significant reduction in the daily system operation cost; moreover the outcomes highlight different demand side response patterns depending on the node of the system where TCLs are connected. In particular, we

- evaluated the value to the security constrained economic dispatch of integrating flexible TCLs in order to minimise the system operational cost,

- demonstrated that TCLs enable turning off expensive peaking generators required to supply the demand when it achieves its maximum level. This action has the additional effect of temporally alleviating the congested transmission line that connects the two nodes of the system considered,
- illustrated how the optimal charging/discharging policy for TCLs may considerably differ if the devices are connected to different bus bars of the system,
- showed that the TCLs' provision of response services can reduce the need to operate several generators part-loaded; as a consequence we demonstrate that the TCLs' flexibility provides an overall system cost reduction and a reduction of CO_2 emissions,
- performed various sensitivity studies varying the flexible TCL penetration and the maximum capacity of the tie line to quantify the system cost savings and the savings obtained by individual appliances. This demonstrates that high per-appliance savings are occur low penetrations of TCLs; this may serve as an incentive to quickly start a new profitable business.

In Chapter 7 we first introduced the improved version of the DSRM for a heterogeneous population of thermostatically controlled loads. Afterwards, the DSRM has been integrated in an advanced scheduling routine to minimize the annual operational cost of the system, optimizing the energy consumption of the appliances and the allocation of frequency services. This chapter provides the final conclusions of this thesis. It is really economically effective integrating a large population of TCLs to optimize the system operation and improve the security of the network. Following the bottom up structure of the sections, we

- built the DSRM in compliance with three characteristics; the first one regards the flexible TCLs´ energy consumption: TCLs can increase (decrease) the consumption in order to (a) perfrom energy arbitrage exploiting the cost differences at different time intervals and (b) provide more (less) response services (following the time-dependent network requirements). The second feature is the full controllability of recovery phase; hence TCLs can absorb an extra amount of power (supplied by reserve generators) to make the energy recovery quicker; the last characteristic is the guaranteed deliverability of the scheduled response regardless of the actual occurrence of generators' failures at previous time step. This represents an

important contribution as it makes the TCLs´ response provision as reliable as the generators' ability to provide frequency control.

- modified the standard formulation of an advanced stochastic unit commitment model to implement the DSRM. Hence, we upgraded the inertia-dependent frequency constraints to calculate fast frequency response requirements,

- compared the scheduling model with the inclusion of DSRM to an alternative model for demand side response; the proposed model shows its effectiveness as it provides higher annual cost savings and higher wind curtailments savings (with respect to the case without flexible TCL control),

- performed sensitivity analysis by varying the maximum response capacity of conventional generators and the settings of the primary response delivery time. The first study shows the need for a simultaneous and comparable provision of primary and secondary response in order to provide high cost savings. The second study confirms the benefits varying the current response provision time of 10s to 5s (in line with NG proposals),

- explored the value of varying the TCLs´ recovery pattern; we identify that the optimal setting for the time of occurrence of the recovery consumption's peak Δt_3 would be an intermediate value ($\Delta t_3{}^* = 10$min if $\Delta t_3 = \{5,10,15\}$); small values are penalised by strict generators' rump rates constraints while, for large values, the TCLs´ consumption is limited by the TCLs´ maximum power threshold and the increased amount of extra reserve required,

- discussed the impact that the use of rolling planning in the scheduling routine has on the TCLs´ quality of the service. Three solutions are compared; the first is not realistic as it does not maintain the average energy (and thus average temperature) over the optimization horizon; both other methods are valid as the average energy is maintained over the optimization horizon. The difference between these methods lies in the strict respect of a constraint on the average energy during the actual physical day; a final assessment should include concerns on the quality of the services requirements and is beyond of the scope of this thesis.

8.2 Further work

At the end of each chapter, we proposed extensions the work presented; some of those improvements have already been taken into consideration within the following chapters. However, there are still aspects that might be considered to improve and continue this work. This thesis has been developed achieving three steps with a logical order: TCL modelling, TCL control and system level impact; hence, future modifications should follow an approach coherent with that order.

It is therefore possible to increase the accuracy of dynamic models of the appliances considered in this work. Afterwards, the impact of the demand side response from TCLs should also include other sectors in addition to the refrigeration one; modelling heating systems such as heat pumps or the refrigeration units located in supermarket will boost the beneficial impact of demand side response as these kind of appliances may offer much larger flexibility compared, for instance, to domestic refrigerators. As a consequence, the aforementioned technologies may be able to simultaneously provide a large contribution of response (and maybe reserve) services and, in the meantime, they would beneficially contribute to the re-shape the system demand level in order to avoid peak-load situations that normally lead to high energy cost. Modelling these devices, especially large refrigeration units, is not a trivial task and the dynamic order of the thermal models could rapidly grow.

Moreover these more complex appliances may not be equipped with only one compressor, therefore intermediate power states have to be used instead of the regular 0 -1 levels. In addition, with regard to the heating systems, these TCLs are not continuously connected to the grid; this aspect introduces a further complication in the respect of the quality of the service. Thus it is worth pointing out that the models presented in this thesis and applied on refrigeration units can be directly extended to heating system. These considerations imply the upgrade of the control strategy in order to design a more sophisticated class of controllers; the developed control strategy may be able to modulate the expected power consumption of each individual unit even between intermediate power levels. Extensions to higher-order devices' models could be permitted together with modifications to enlarge the range of the instantaneous power levels.

Improvements can also be achieved with regard to the unit commitment model in Chapter 7. This model does not currently incentivize the faster provision of primary response service as it is not able to recognize the different ramp rates of the generators. Fast responding generators and

demand side response actors, if properly incentivized, could provide significant cost savings and wind curtailment reduction, thus lowering the expensive fast frequency regulation requirements. A future SUC model should also include power flow constraints that are neglected in this first implementation. Moreover, under certain limits, it will be beneficial to make the TCLs services' commitment flexible; for instance, at each time step, the energy recovery pattern could vary in accordance with the ramp rate constraints of the reserve generators available at that time. The value for TCLs providing ancillary service to the system has to include the investment costs required to make several millions of devices actually able to provide system support.

There are interesting research cues regarding the demand response with TCLs that have not been touched in this thesis although the tools provided in these chapters (modelling and control) would represent an important starting point. In effect, the control of thermostatic loads by means of the proposed decentralised controlled should be applied to a more realistic network model to give a more realistic measure of the impact of demand side response. In a large bus-bar system the frequency evolution after a generator outage would not be the same at each node, therefore it would be relevant to investigate the effect of not-simultaneous activations of demand side power reductions on the frequency evolution. In this thesis the effect of the reactive power reduction associated to the TCLs load shedding has been ignored; further studies on a more realistic network will investigate the effect of such a distributed load shedding on the voltages profiles. Finally the smart control of TCLs could be included in advanced intertripping schemes to increase the transfer capacity of selected transmission lines in presence of large shares of RES. Due to stability constraints this further development would require very a fast activation of demand side actions.

References

[1] United Nations Environment Programme, «Intergovernal Panel on Climate Change (IPCC),» [Online]. Available: http://www.ipcc.ch/index.htm.

[2] United Nations, «Kyoto Protocol to the United Nnations Framework Convention on Climate Change,» 1998. [Online]. Available: http://unfccc.int/resource/docs/convkp/kpeng.pdf.

[3] United Nations - Framework Convention on Climate Change, «COP 16 - Cancun Agreements,» [Online]. Available: http://unfccc.int/meetings/cancun_nov_2010/session/6254/php/view/reports.php.

[4] United Nations - COP 21, [Online]. Available: http://climate-l.iisd.org/events/unfccc-cop-21/.

[5] Report from the commission to the European Parliament and the Council , «Progress towards Chieving the Kyoto and EU 2020 Objectives,» 2014. [Online]. Available: http://ec.europa.eu/clima/policies/g-gas/docs/com_2014_689_en.pdf.

[6] European Commission, «Europe 2020,» [Online]. Available: http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/targets/index_en.htm.

[7] European Commission, «Climate Action - Roadmap for 2050 emission reduction targets,» [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011DC0112.

[8] Committee on Climate Change , «Building a low-carbon economy – the UK's contribution to tackling climate change,» [Online]. Available:

http://archive.theccc.org.uk/aws3/key%20messages%20-%2012%20page%20version.pdf.

[9] UK Department of Energy and Climate Change, «2013 UK Greenhouse Gas Emissions Provisional,» [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/295968/20140327_2013_UK_Greenhouse_Gas_Emissions_Provisional_Figures.pdf.

[10] European Commission, «EU Directive 2009/28/EC,» 2008. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32009L0028:EN:.

[11] National Grid, «UK Future enrgy scenarios,» [Online]. Available: http://www2.nationalgrid.com/uk/industry-information/future-of-energy/future-energy-scenarios/.

[12] National Grid, «Frequency response technical sub group report,» Nov. 2011. [Online]. Available: http://www.nationalgrid.com/uk/electricity/codes/gridcode/workinggroups/freqresptsg/.

[13] RenewableUK, «UK Wind Energy Database (UKWED),» [Online]. Available: http://www.renewableuk.com/en/renewable-energy/wind-energy/uk-wind-energy-database/.

[14] The Wind Power - wind turbines and wind farms database, «Countries statistics,» [Online]. Available: http://www.thewindpower.net/statistics_countries_en.php.

[15] GL Garrad Hassan, «UK Generation and Gemand Scenarios for 2030,» [Online]. Available: http://assets.wwf.org.uk/downloads/positive_energy_glgh_technical_report.pdf.

[16] HR Governament , «2050 Pathway Analysis,» July 2010. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/42562/216-2050-pathways-analysis-report.pdf.

[17] Digest of United Kingdom energy statistics , «Renewable sources of energy: chapter 6,» [Online]. Available: Renewable sources of energy: chapter 6, Digest of United Kingdom energy statistics (DUKES).

[18] G. Delille, B. Francois and G. Malarange, «Dynamic frequency control support by energy sotrage to refuce the impact of wind and solar generation on isolated power system's inertia,» *IEEE Transactions on Sustainable Energy,* vol. 3, n. 4, pp. 931-939, 2012.

[19] A. Mullane ans M. O'Malley, «The inertial response of induction machine based wind turbines,» *IEEE Trans. Power Syst.,* vol. 20, n. 3, pp. 1496 - 1503, 2005.

[20] National Grid, «GB Security and Quality of Supply Standard,» [Online]. Available: http://www.nationalgrid.com/NR/rdonlyres/FBB211AF-D4AA-45D0-9224-7BB87DE366C1/15460/GB_SQSS_V1.pdf.

[21] National Grid, «Report of the investigation into the automatic demand disconnection following multiple generation losses and the demand control response that occurred on the 27th May 2008,» [Online]. Available: http://www2.nationalgrid.com/assets/0/745/746/3464/3466/3488/3475/623fcdf6-31e3-416a-bfb7-c4edbfd763d2.pdf.

[22] European Technology Platform SmartGrids, «Strategic Deployment Document for Europe's Electricity Networks of the Future,» 2010. [Online]. Available: http://www.smartgrids.eu/documents/SmartGrids_SDD_FINAL_APRIL2010.pdf.

[23] B. Silva, C.L. Moreira, L. Seca, Y. Phulpin and J.A. Peas Lopes, «Provision of Inertial and Primary Frequency Control Services Using Offshore Multiterminal HVDC Networks,» *IEEE Transactions on Sustainable Energy,* vol. 3, n. 4, pp. 800 - 808, 2012.

[24] Y. Ota, H. Taniguchi, T. Nakajima, K.M. Liyanage, K. Shimizu, T. Masuta, J. Baba and A. Yokoyama, «Effect of autonomous distributed vehicle-to-grid (V2G) on power system frequency control,» in *International Conference on industrial and Information Systems (ICIIS)*, Mangalore, 2010.

[25] M.A. Ortega-Vazquez, «Optimal scheduling of electric vehicle charging and vehicle-to-grid services at household level including battery degradation and price uncertainty,» *IET Generation, Transmission & Distribution,* vol. 8, n. 6, pp. 1007 - 1016, 2014.

[26] R.Moreno, R.Moreira and G.Strbac, «A MILP model for optimising multi-service portfolios of distributed energy storage,» *Apllied Energy - Elsevier,* 2014.

[27] G. Lalor, A. Mullane, and M. O'Malley, «Frequency Control and Wind Turbine Technologies,» *IEEE Trans. Power Syst.,* vol. 20, n. 4, Nov. 2005.

[28] M. Aunedi, E.O. Calderon, V. Silva, P. Mitcheson and G. Strbac , «Economic and environmental impact of dynamic demand,» 2008. [Online]. Available: www.supergen-networks.org.uk/50/file.pdf.

[29] J. Short, D. Infield, and L.L. Freris , «Stabilization of Grid Frequency through Dynamic Demand Control,» *IEEE Trans. Power Syst.,* vol. 22, n. 2, Aug. 2007.

[30] Department of energy and climate change, «Digest of UK energy statistics,» 2009.

[31] D. Angeli, and P.A.Kountouriotis, «A Stochastic Approach to "Dynamic Demand" Refrigerator Control,» *IEEE Trans. Control Syst. Techn.,* vol. 20, n. 3, May 2012.

[32] M. Stadler, W. Krause, M. Sonnenschein and U. Vogel , «Modelling and evaluation of control schemes for enhancing load shift of electricity demand for cooling devices,» *Elevier Enviromental Modeling and Software ,* vol. 24, pp. 285-295, 2009.

[33] C.Perfumo, E. Kofman, J.H. Braslavsky and J. Ward, «Load management:Model-based control for population of thermostatically controlled loads,» *Energy Conversion and Management,* vol. 55, pp. 36-48, Mar. 2012.

[34] J.L. Mathieu, S.Koch and D.S.Callaway, «State Estimation and Control of Electric Loads to Manage Real-Time Energy Imbalance,» *IEEE Trans. Power Syst,* vol. 28, n. 1, pp. 430-440, 2013.

[35] T. Keep, F. Sifuentes, D. Auslander and D. Callaway, «Using load switches to control aggregated electricity demand for load following and regulation,» in *IEEE Power and Energy Society General Meeting,*, San Diego, 2011.

[36] C. Sonntag, A. Devanathan, S. Engell, S and O. Stursberg, «Hybrid Nonlinear Model-Predictive Control of a Supermarket Refrigeration System,» in *IEEE International Conference on Control Applications CCA*, Singapore, 2007.

[37] S. Shafiei, H. Rasmussen and J. Stoustrup, «Modeling Supermarket Refrigeration Systems for Demand-Side Management,» *Energies,* vol. 6, n. 2, 2013.

[38] http://web.ornl.gov/~webworks/cppr/y2002/pres/113141.pdf, «Advances in supermarket refrigeration systems,» [Online]. Available: http://web.ornl.gov/~webworks/cppr/y2002/pres/113141.pdf.

[39] F. Schweppe, «Frequency adaptive, power-energy re-scheduler». United States Brevetto US4317049 A, Sept. 1979.

[40] S. Ihara and F. Schweppe, «Phisically based modeling of cold load pickup,» *IEEE Transactions on Power Apparatus and Systems,* Vol. %1 di %2PAD-100, n. 9, pp. 4142-4150, 2013.

[41] D. Callaway, «Tapping the energy storage potential in electric loads to deliver load followingand regulation, with appliacation to wind energy,» *Energy Conversion and Management,* vol. 50, n. 5, pp. 1389-1400, 2009.

[42] S. Kundu, N, Sinitsyn, S. Backhaus and I.Hiskens, «Modeling and control of thermostaticaly controlled loads,» in *17th Power Systems Computation Confernce*, Stockholm, 2011.

[43] Y.Tan and D.Kirschen, «Co-optimization of Energy and Reserve in Electricity Markets with Demand-side Participation in Reserve Services,» in *PSCE*, Atlanta, 2006.

[44] J.Wang N.Redondo and F.Galiana, «Demand-side reserve offers in joint energy/reserve electricity markets,» *IEEE Trans. Power Syst.,* vol. 18, n. 4, pp. 1300-1306, Nov. 2003.

[45] E.Karangelos and F. Bouffard, «Towards Full Integration of Demand-Side Resources in Joint Forward Energy/Reserve Electricity Markets,» *IEEE Trans. Power Syst.,* vol. 27, n. 1, pp. 280-289, Feb. 2012.

[46] M. Aunedi, P. Kountouriotis, J.Calderon, D.Angeli and G.Strbac, , «Economic and Enviromental Benefits of Dynamic Demand in Providing Frequency Regulation,» *EEE Transactions on Smart Grids,* vol. 4, n. 4, pp. 2036 - 2048, 2013.

[47] J. Mathieu, M. Kamgarpour, J.Lygeros, and D. Callaway, «Energy Arbitrage with Thermostatically Controlled Loads,» in *European Control Conference (ECC)*, Zurich, 2013.

[48] D.Papadaskalopoulos, G. Strbac, P. Mancarella, M. Aunedi and V. Stanojevic, «Decentralized Participation of Flexible Demand in Electricity Markets—Part II: Application With Electric Vehicles and Heat Pump Systems,» *Transactions on Power Systems,* vol. 28, n. 4, pp. 3667-3674, Nov 2013.

[49] D.S. Callaway and I.A. Hiskens , «Achiving controllability of electric loads,» *Proceedings of the IEEE,* vol. 99, n. 1, pp. 184-199, 2011.

[50] National Grid , «Frequency control by demand managment,» [Online]. Available: http://www2.nationalgrid.com/uk/services/balancing-services/frequency-response/frequency-control-by-demand-management/.

[51] National Grid, «Demand Side Balancing Reserve (DSBR),» [Online]. Available: http://www2.nationalgrid.com/mediacentral/uk-press-releases/2014/national-grid-to-contract-for-new-balancing-services/.

[52] PJM Interconnection LLC, [Online]. Available: http://www.pjm.com/..

[53] NYISO, [Online]. Available: http://www.nyiso.com/public/index.jsp.

[54] Hydro Quebec, [Online]. Available: http://www.hydroquebec.com/..

[55] ENTSO-E, «Demand Connection Code,» 2013. [Online]. Available: https://www.entsoe.eu/major-projects/network-code-development/demand-connection/Pages/default.aspx#documents.

[56] S. Tindemans, V. Trovato and G. Strbac, «Decentralised control of thermostatic loads for flexible demand response,» *IEEE Transactions on Control System Technolgy,* 2014 - in press.

[57] F. Teng, V. Trovato and G.Strbac, «Stochastic scheduling with inertia-dependent frequency regulation,» *submitted to IEEE Transactions on Power Systems,* 2014.

[58] P. Kundur, Power system stability and control, London: McGraw-Hill,, 1994, .

[59] Ofgem, «National Electricity Transmission System Security and Quality of Supply Standard (NETS SQSS): Review of Infeed Losses (GSR007 as revised by GSR007-1),» [Online]. Available: https://www.ofgem.gov.uk/ofgem-publications/52793/gsr007-decision-letter-final.pdf.

[60] Wikipedia , «Nuclear power in the United Kingdom,» [Online]. Available: http://en.wikipedia.org/wiki/Nuclear_power_in_the_United_Kingdom.

[61] G. Tarnowski, P.Kjaer, S. Dalsgaard, and A. Nyborg, , «Regulation and Frequency Response Service Capability of Modern Wind Power Plants,,» in *IEEE Power and Energy Society General Meeting*, Minneapolis, 2010.

[62] L. Holdsworth, J. B. Ekanayake and N. Jenkins, «Power system frequency response from fixed speed and doubly fed induction generator-based wind turbines,» *Wind Energy,* vol. 7, n. 1, pp. 21-32, Mar. 2004.

[63] N, Ullah, T. Thiringer and D. Karlsson, «Temporary primary frequency control support by variable speed wind turbines - potentials and applications,» *IEEE Trans. Power Syst.,* vol. 23, n. 2, pp. 601-612, May 2008.

[64] L. Wu, and D. Infield , «Modelling the provision of Inertial Response from Variable Speed Wind Turbines,» in *Renewable Power generation, IET Conference*, 2011.

[65] National Renewable Energy Laboratory, «Cost and performance data for power generation technologies,» Feb. 2012. [Online]. Available: http://bv.com/docs/reports-studies/nrel-cost-report.pdf.

[66] R. Marconato, Electric power systems vol.2, Milan: CEI- Italian Electrothecnical Committe , 2004.

[67] European Commission, «Intelligent Energy - Europe' Programme,» [Online]. Available: http://www.smart-a.org/.

[68] Department for business enterprise and regulatory reform, «Dynamic Demand - Governament response to clause 18 of the climate change and sustinable energy act,» 2007. [Online]. Available: http://www.theapplicationhome.com/NewFiles/DynamicDemand-BERR.pdf.

[69] Pennsylvania New Jersey Maryland Interconnection - PJM, «Demand side response,» [Online]. Available: http://www.pjm.com/Globals/Training/Courses/ip-dsr.aspx.

[70] Y.A. Cengel, Introduction to thermodynamics and heat transfer, United States: McGraw-Hill, 2009.

[71] Whirpool, «Technical datasheets of cold appliances,» [Online]. Available: www.whirlpool.co.uk.

[72] Williams Refrigeration, «Commercial appliances,» [Online]. Available: www.williamsrefrigeration.co.uk.

[73] Navigant Consulting, Inc., «nergy Savings Potential and R&D Opportunities for Commercial Refrigeration,» 2009.

[74] Mathworks, « Matlab, the language of thechnical computing,» [Online]. Available: http://www.mathworks.co.uk/products/matlab/.

[75] M. Kamgarpour et al., «Modeling options for demand side participation of thermostatically controlled loads,» in *IREP Symposium*, Crete, 2013.

[76] National Grid, «High frequency response service,» [Online]. Available: http://www2.nationalgrid.com/uk/services/balancing-services/frequency-response/mandatory-frequency-response/.

[77] S.Pourmousavi and M.Nehrir, «Introducing Dynamic Demand Response in the LFC Model,» *IEEE Tansactions on Power System,* vol. 29, n. 4, 2014. .

[78] National Grid, «Frequency response service summary,» [Online]. Available: http://www2.nationalgrid.com/UK/Services/Balancing-services/Frequency-response/Mandatory-Frequency-Response/.

[79] National Grid, [Online]. Available: http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-explorer/.

[80] National Grid, «Outcome - Energy services - Market information,» [Online]. Available: http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-explorer/Outcome-Energy-Services/.

[81] R. Bharathi, M.J. Kumar, D. Sunitha and S. Premalatha., «Optimization of combined economic and emission dispatch problem — A comparative study,» in *Power Engineering Conference, IPEC*, Singapore, 2007.

[82] G. Tao Guo, M.I. Henwood and M. van Ooijen, «An algorithm for combined heat and power economic dispatch,» *IEEE Transactions on Power Systems,* vol. 11, n. 4, pp. 1778 - 1784, 1996.

[83] Statistic Brain, [Online]. Available: http://www.statisticbrain.com/hotel.

[84] [Online]. Available: http://www.beerandpub.com/statistics.

[85] [Online]. Available: http://www.telegraph.co.uk/finance/newsbysector/retailandconsumer/coffee.

[86] [Online]. Available: http://ratings.food.gov.uk/search-a-local-authority-area.

[87] [Online]. Available: http://www.voa.gov.uk/corporate/Publications.

[88] National Grid, «Electricity Ten Years Statement,» 2013. [Online]. Available: http://www2.nationalgrid.com/UK/Industry-information/Future-of-Energy/Electricity-ten-year-statement/.

[89] «Cost and performance data for power generation technologies,» 2012.

[90] «Carbon footprint of electric generation,» 2006.

[91] Ofgem, «D09 OC6.6 Automatic Low Frequency Demand Disconnection,» [Online]. Available: https://www.ofgem.gov.uk/publications-and-updates/d09-oc6.6-automatic-low-frequency-demand-disconnection-lfdd.

[92] C. Jianming, F. Lee, A. Breipohl and R. Adapa, «Scheduling direct load control to minimize system operation cost,» *IEEE Transactions on Power Systems,* vol. 10, n. 4, 1995.

[93] I.A.Erinmez, D.O. Bickers, G.F. Wood and W.W. Hung, «NGC experience with frequency control in England and Wales-provision of frequency response by generators,» in *IEEE Power Engineering Society - Winter Meeting*, New York, NY, 1999.

[94] H. Chavez, R. Baldick and S.Sharma, «Governor Rate-Constrained OPF for Primary Frequency Control Adequacy,» *IEEE trans. Power Syst.,* vol. 29, n. 3, pp. 1473-1480, 2014.

[95] Y. Lee and R. Baldick, « A Frequency-Constrained Stochastic Economic Dispatch Model,» *IEEE Trans. Power Syst.,* vol. 28, n. 3, pp. 2301-2312, 2013.

[96] Energy Networks Association and National Grid, «Frequency Changes during Large Disturbances and their Impact on the Total System,» 2013.

[97] W. Qianfan, J-P. Watson and Y. Guan, «Two-Stage Robust Optimization for-Contingency-Constrained Unit Commitment,» *IEEE Trans. Power Syst.,* vol. 28 , n. 3, pp. 2366-2375, 2013.

[98] Department of Energy & Climate Change, «Digest of UK energy statistics,» [Online]. Available: https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes.