Les Cahiers du GERAD ISSN: 0711–2440

Operations research approaches for building demand response in a smart grid

M. F. Anjos,
J. A. Gomez

G–2017–55

June 2017

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Operations research approaches for building demand response in a smart grid

Juan A. Gomez
Miguel F. Anjos

GERAD & Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7

juan.gomez@polymtl.ca
miguel-f.anjos@polymtl.ca

June 2017

Les Cahiers du GERAD

G–2017–55

Abstract: Electric power systems need to ensure that production and demand of electricity are continuously in balance. With fundamental changes taking place in the power grids of many countries due to a variety of technological and policy developments, there is a need to obtain additional flexibility to achieve this essential power balance. Demand response refers to the collection of all the means to obtain this flexibility from the demand side of the balance. We present a selection of contributions of operations research to the provision of demand response by the residential, commercial and institutional sectors of the economy. The aspects covered include electricity tariffs, building energy management systems, load estimation, local generation, electric vehicles, energy storage, and building-level aggregation. We conclude with a brief discussion of current opportunities for operations research to support the development and realization of the potential of demand response.

Keywords: Smart buildings, demand response, residential sector, energy management systems, classification, integer programming, stochastic programming, bi-objective optimization

1 Introduction

Electricity is a critical resource for society. The ubiquitousness of electrical devices in our living environment makes it an essential part of our daily life, although it is often taken for granted due to the great success of our contemporary electric power systems. These systems were built on the principle of bulk production of electricity in a limited number of locations (to achieve economies of scale) coupled with a large-scale transmission and distribution grid to bring the electricity to the consumers.

Since the 1990s, electric power systems in many jurisdictions around the world have started undergoing significant changes. The specifics vary between jurisdictions, but some fundamental changes are ubiquitous and brought about by global technological and policy developments. These developments include:

- The increasing sophistication of *data-gathering devices* such as smart meters, voltage sensors, and fault detectors, together with the two-way communication technology to transmit the data automatically, offers power distribution companies the ability to manage their operation more closely. While the extent to which the power grid has been automated and computerized varies from one jurisdiction to another, this is an irreversible trend in the industry.
- In many jurisdictions there has been an increase in the number of occurrences when there is limited (or no) spare power generation capability when compared with the demand for electricity, resulting in a tightly constrained operating context. The moments of high power consumption levels are called *demand peaks* and they are a major concern for the system operator. While such extreme situations currently occur only during a few days of the peak season (e.g., winter in Quebec or summer in New York), the trend is towards having overall tighter operating margins and higher capacity factors, meaning that the system will be operating more often near its maximum possible production level. Indeed in some systems, the growth in peak demand surpasses the growth in annual demand for electricity.
- The ever-increasing amount of electricity being generated from *renewable sources*. Much of this energy is scattered in a large number of locations throughout the grid, with relatively small quantities produced at each location. In addition to the distributed nature of this generation, its intermittent nature, notably from wind power and solar power, leads to major technical challenges to ensure the stability and reliability of the grid.
- Progress in *energy storage* technologies has made it technically feasible to store energy in increasingly large quantities. Grid-scale energy storage can fulfill a number of important functions from the grid operator's perspective, such as smoothing the output fluctuations from intermittent renewable generation, and supporting the grid operation during load peak periods by releasing energy stored at other times. On the other hand, the prospect of combining renewable generation with storage at the consumer level throughout the system brings about concerns about the economic operation of the grid, and leads to the need for tariffs to account for both the provision of energy and of a connection to the grid.
- The increasing penetration of electric vehicles will change the nature and the magnitude of the load on the grid. More generally, *electric transportation* will become a major component of the load in future power grids. On the other hand, their batteries are effectively energy storage devices and can thus in principle (though the practice is complicated) fulfill some of the operational functions of storage that were mentioned above.

These developments require a paradigm shift in the planning, design and operation of power systems; this shift is referred to as *smart grid*. A smart grid is the combination of a traditional electrical power production, transmission and distribution system with information and energy both flowing back and forth between suppliers and consumers. This combination is expected to deliver energy savings, cost reductions, and increased reliability and security.

Every power system needs to ensure that production and demand of electricity are continuously, and almost perfectly, in balance. Traditional power systems achieve this by adjusting production to match demand. This is possible so long as enough of the output of generators can be adjusted as needed by the system. However, with the increasing penetration of renewable generation, the proportion of generation that can be controlled decreases. Indeed this intermittent generation is normally accounted for on the demand side, meaning that the system needs to meet the *net demand*, which is equal to the total electric demand in the system minus the contribution of intermittent generation. Notwithstanding the progress in energy storage capabilities, this new reality creates a need for the power grid to obtain sufficient *flexibility* from the real demand to achieve the essential power balance. This means not only reducing peaks in consumption but, more generally, reducing the fluctuation of net demand. The collection of approaches available to obtain this flexibility from the demand side of the balance is commonly referred to as *demand response (DR)*.

The economic potential for DR is important. The DR market in the USA alone generated in 2011 approximately USD$6 billion in direct revenues for DR providers, even though DR programs have until now focused mainly on industrial and some large commercial consumers [34]. Thus the potential for residential, commercial and institutional buildings to provide DR, still largely untapped, could have significant value for consumers.

Worldwide, the power consumption of buildings accounts for an estimated 40% of global energy consumption [44]. In the United States, residential and commercial buildings represent around 70% of the total energy demand and the potential of peak DR was estimated in 2015 at 8.7 GW in the United States [4]. In Canada, space heating is responsible for more than 60% of the total residential energy consumption, due to the cold climate [2], and electric baseboards account for 27% of heating nationally, and for 66% in the (hydro-rich) province of Quebec. Furthermore, the province of Ontario is a summer-peaking system with a high penetration of air-conditioning systems [1, 3].

This potential is however challenging to realize in practice. A 2012 survey by the US Federal Energy Regulatory Commission (FERC) reports on barriers to wider adoption of DR [5]. These include policy issues such as a lack of standards for communicating DR pricing and insufficient consumer engagement, but they also include technical limitations such as the need to improve measurement and verification of demand reductions, and a lack of DR forecasting and estimation tools.

This tutorial is concerned with the opportunities for operations research (OR) to support the provision of DR by the residential, commercial and institutional sectors of the economy. This is a realistic prospect with the advent of *smart buildings* that have automated management systems with the capability to learn (and even anticipate) the demand for light, temperature, cooking and other services, and hence to support energy management strategies [36]. This is an active area of research and thus there is a large amount of relevant literature. For the sake of focusing on the major issues and challenges, and also for the sake of brevity, we do not provide a survey of the literature. Instead we focus on selected recent developments to highlight the opportunities for OR-based approaches to address the challenges in the building-based provision of DR. For ease of presentation, we focus the presentation on the residential sector, but most of the techniques discussed can be adapted to the commercial and institutional contexts as well. Moreover we focus the discussion on the provision of DR as a means to smooth the demand curve, but we note that DR is increasingly able to provide so-called ancillary services that are essential for the secure operation of the power system and have been typically provided by generators.

This tutorial is structured as follows. In Section 2 we address the question of how to set electricity tariffs for residential customers so as to support the provision of DR. In particular, the recently proposed concept of a Time-and-Level-Of-Use tariff is considered in some detail in Section 2.4. Section 3 looks at the use of energy management systems (EMS) to coordinate the consumption of energy in a home or building. Section 4 examines the question of estimating the power capacity profiles for a building for the purpose of supporting the implementation of a Time-and-Level-Of-Use tariff. Specifically Section 4.1 considers loads of a thermal nature, and Section 4.2 considers major loads that offer some flexiblity in scheduling. Section 5 presents an optimization model to support the integration of electric vehicles and solar panels in a residential building. The model include scheduling of loads and charging of the EVs so as to miminize the amount of energy procured from the grid. Section 6 is concerned with an optimization model to support the aggregation of residential DR potentials to make their use more effective in practice. Section 7 concludes the tutorial with some challenges for OR in the area of building DR.

Because the modelling of DR involves different time scales, we use the following terminology: time steps (duration of a few minutes) at the operational level, time frames (duration of one hour) at the planning level, and time windows (duration of several hours) at the level of the application of tariffs. This convention is applied consistently throughout this tutorial. At the same time, the diversity of topics makes it difficult to ensure a consistent notation throughout the tutorial. For this reason, the notation used is defined in each section.

2 Tariffs for the provision of DR

The success of DR initiatives critically depends on the setting of suitable pricing policies that provide sufficient financial rewards for the consumers to participate. The importance of this economic motivation is explicitly mentioned in FERC's definition of DR:

> "Changes in electric use by demand-side resources from their normal consumption patterns in response to changes in the price of electricity, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized" [5].

We discuss in this section the question of how to set the price of electricity to support the provision of DR. The setting of incentive payments is not addressed here.

2.1 Electricity costs and the residential context

At the system level, there are several components to the price of electricity. First, there is the cost of energy itself. In jusrisdictions where there is an electricity market, this will be the marginal cost of energy as determined by the equilibirum between supply and demand. In the absence of a market structure, this cost is regulated on the basis of generation cost (including infrastructure costs that are amortized in the long-term). Other factors are related to the transmission of electricity, such as the costs of line losses and of congestion on the grid. The use of locational marginal prices bundles all these costs into an overall price that varies across the grid.

The complexity of electricity pricing normally does not directly affect residential customers who pay for their consumption based on the time-invariant tariffs set by the electricity provider of their choice, and these tariffs are usually fixed for a certain duration. From the point of view of residential DR, the situation is therefore that direct reaction to the evolution of market prices is mostly not happening at present, even in the presence of smart meters. While this creates in principle an opportunity for residential EMSs with the ability to react rapidly to the evolution of prices and change the consumption of electricity accordingly, as per the above definition, there remains the question of how willing customers would be to follow the market signals so closely. Indeed, while customers are open to dynamic pricing policies, they prefer simple programs to complex ones, and frameworks that support automatic responses without direct intervention or monitoring of electricity prices by the consumers are seen as a prerequisite for successful participation of residential customers in DR programs [15]. Customers are thus likely to prefer to pay more for the convenience of time-invariant tariffs, and thus not participate in the provision of DR (unless the financial incentives are particularly high, which does not seem realistic except in uncommon circumstances and in the presence of a CPP surcharge, as described below).

2.2 Time Of Use (TOU) tariffs

A common pricing scheme that encourages some provision of DR while maintaining tariff stability is Time Of Use (TOU). A TOU tariff divides a day into time windows, and specifies different prices for those time windows using at least two price levels: a lower price for the off-peak windows and a higher price for those on-peak. In some jurisdictions an intermediary mid-peak price is also used. The time windows and prices are set in advance and do not change, except possibly according to the seasons of the year. For example, Ontario has a TOU tariff for winter (November-April) and another one for summer (May-October). The idea is that a TOU tariff reflects the expected conditions of the grid, and hence roughly approximate the market price. This scheme is already available in a number of jurisdictions, both in North America (e.g. Ontario and Massachusetts) and in Europe (e.g. France and Italy).

Other schemes that attempt to approximate the market price include Critical Peak Pricing (CPP) and Variable Peak Pricing (VPP). CPP tariffs augment a time-invariant or TOU tariff with a critical (high) surcharge during time windows when specific, pre-defined conditions occur in the grid and/or the market. The time windows may be fixed in advance, or they may vary depending on the system needs. In the latter case, customers are notified a few hours in advance, and can respond accordingly. The impact of CPP at the residential level was demonstrated in [22]. VPP tariffs are a variation of TOU in which the different pricing periods are defined in advance but the price for on-peak periods varies, again depending on the grid and/or market conditions.

2.3 Multi-TOU tariffs

One of the practical concerns in implementing DR is the risk of creating *rebound peaks* in the aggregate demand [35]. Such peaks may arise if a large number of smart building EMSs respond by shifting demand from peak times to a given non-peak period, and thereby create a (possibly larger) peak during the latter period. Even if not so extreme, this response can lead to increased load fluctuations on the grid, see e.g. [27], hence going against the objective of reducing fluctuations in net demand.

The concept of a *Multi-TOU* tariff is proposed in [34] as a means to resolve this difficulty. A Multi-TOU tariff is essentially a group-based TOU tariff in which the customers are divided into groups, and each group is assigned a slightly different TOU tariff. For instance, in the case of two groups, the moment at which the price changes from off-peak to on-peak may differ by one hour between the two groups. The result is that every customer has a TOU tariff but the undesirable effects of synchronized behaviour are reduced. Issues of fairness arising from the use of slightly different tariffs can be addressed by rotating the customers among the groups. With this policy every customer has the same overall combination of tariffs in the course of a given period of time (e.g. one year). A Multi-CPP tariff can be set up in a similar fashion.

2.4 Time-and-Level-Of-Use (TLOU) tariffs

All the aforementioned tariffs aim to adjust the price of electricity, but do not directly deal with the level of power used by the customers. However, the interest for the possibility of pre-specifying a level of power consumption for a given period of time (as opposed to the total energy consumption over that period) is likely to increase in the coming years for at least two reasons.

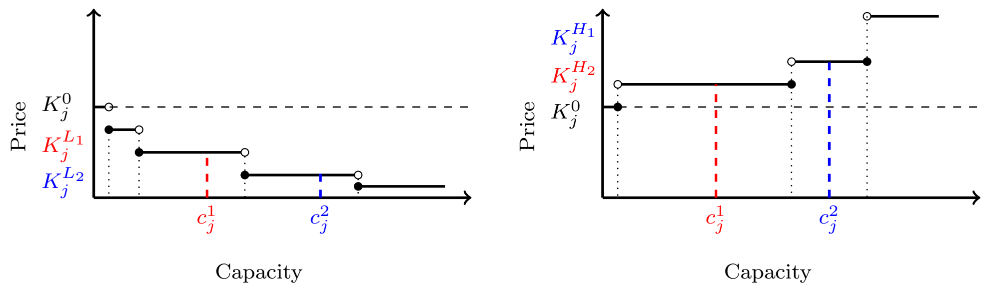
First, as mentioned above, the advent of the smart grid increases the importance of reducing the fluctuations of net demand, and a direct means to provide DR in terms of the level of power is an attractive possibility from the operational perspective.

Second, while an increase in use of distributed generation and of energy storage technologies makes it likely that the business of providing energy will grow more slowly (or possibly even shrink), it will create opportunities for the business of providing backup electricity services. The business model for such services will replace some the revenue from selling energy with a charge for the provision of backup; in other words, the connection to the grid will itself become a service with an explicit cost greater than that currently bundled into the prices of energy.

Because the latter is directly related to the overall power level in the grid, which in turn directly affects the operating conditions of the grid, another alternative is to consider a tariff that account for both time and level of use. This is the motivation for the Time-and-Level-Of-Use (TLOU) tariff recently proposed in [19]. This is an extension of TOU that specifies a capacity level for each of the time windows of a TOU tariff. The customers then pays a lower price for power consumption up to the specified capacity, and a higher price beyond that threshold. As per the TOU concept, the lower and higher prices vary between time windows.

An important question is the setting of the capacity level for each time window. Because the objective is to support customer participation in DR, the capacity level needs to be set according the customer's needs, yet at the same time, the objective is to keep the individual limits low so that the aggregate of all the customers' limits will allow the grid to plan its operations with a lower level of expected load.

For this purpose, an extended TLOU is proposed in [20] where the customer chooses the capacity level based on a set of possible tariffs and capacities proposed by the electricity provider. The structure of these tariffs is depicted in Figure 1 where the subscript j denotes a time frame.



(a) TLOU tariff up to selected capacity **(b) TLOU tariff beyond selected capacity**

Figure 1: Structure of TLOU tariff options [20]

The default is the TOU tariff K_j^0 corresponding to the dotted line. For two possible capacity choices $c_j^2 > c_j^1$, we have a lower price $K_j^{L_2}$ in the case of c_j^2 for consumption up to that capacity, but a higher price $K_j^{H_2}$ for consumption beyond it.

Thus the lower and higher prices depend on the capacity level chosen by the customer, and the principle is that as the capacity selected increases, the lower price decreases and the higher price increases. The proposed scheme also includes a booking fee K_j^F per unit of power that is booked in advance by the consumer. Selecting the capacity level is thus a tradeoff: booking a higher capacity c_j^2 results in a lower $K_j^{L_2}$ but a more expensive $K_j^{H_2}$ and a higher total booking cost, while booking a lower capacity means a higher $K_j^{L_2}$ but a lower $K_j^{H_2}$ and a lower total booking cost.

The overall effect is to motivate the customer to aim for a capacity level high enough to meet the household's demand but not higher, and then to organize the household's activities so as to stay within the chosen capacity and avoid consumption levels incurring the higher price. The management of the consumption within this capacity limit is then managed by an EMS, which is the subject of the next section.

3 Energy management systems

We consider a smart building within which energy consumption is coordinated by an energy management system. There is a large and growing body of literature on this subject. Beaudin and Zareipour [10] present a comprehensive review of EMSs focusing on the modeling features and the complexity that arises from the parameter definition in a load scheduling context.

The fundamental purpose of an EMS is to manage the operation of loads inside the building in order to achieve one or more pre-specified objectives in terms of energy and/or power consumption. More generally, an EMS coordinates the interactions of the power-consuming devices in the building, supports the integration of local generation (such as roof-top solar panels), and manages the interaction with the power grid.

These three elements are key for the use of the DR potential in smart buildings. A DR-effective EMS must consider the consumer preferences, the grid requirements and the economic viability for both parties. This challenge can be overcome through the implementation of OR tools.

We suppose that the smart building is willing to provide DR by adjusting its maximum power consumption. A key idea here is the use of *power capacity limits* to enforce the maximum levels of power consumption that the EMS commits to respect during given periods of time. For a sufficiently large number of smart buildings, the aggregate impact of these limits is a reduction on peak demand. The question of effectively aggregating DR resources is addressed in Section 6. We next present an EMS architecture that supports the use of such power capacity limits.

3.1 A Layered EMS architecture

We present in this section the principles of the system architecture proposed in [14] to manage the loads in smart buildings.

One of the issues to address is that a system of loads of different magnitudes (such as a building) is characterized by *multiple time-scales*. Indeed load control is carried out in real-time, on a scale of at most a few minutes, whereas handling price bidding and scheduling of loads are performed in a longer time-scale, typically hours. The architecture uses three layers to handle the aspects corresponding to different time-scales:

- In the bottom layer, the Admission Controller (AC) performs real-time load control. Whenever a request to run a load arrives, the request is accepted if adding it to the current set of loads will still respect the capacity limit; otherwise the request is deferred. Accepted requests can start running right away whereas deferred requests are passed on to the load balancer.
- In the middle layer, the Load Balancer (LB) receives the deferred requests and works out a schedule for them within a given scheduling horizon while taking into account the operational constraints of each load and the capacity limits prescribed by the demand response manager (described next). Once the scheduling is done, the LB communicates to each load the time for it to again make a request to the AC. This next request may be accepted or it may be again deferred, depending on the circumstances at that time.
- In the top layer, the Demand Response Manager (DRM) handles the interaction between the smart building and the grid. The LB provides it with basic performance metrics for the building, such as capacity utilization and rejection rate. It receives cost information from the grid, passing that information to the LB. It also uses information from the load forecasting module. Using all this information, the DRM sets the capacity limit for each time frame in the planning horizon.

Let us now consider each of the layers in turn.

3.1.1 Admission controller

The AC is the entry point for each load to access power, and the main requirement here is to make a quick decision to accept or defer a request. The algorithm presented in [14] works as follows. First, the algorithm sorts all the requests from the highest to the lowest priority value, where the priority value is a measure of the urgency of the request. For instance in [21], the priority value for a room's heating/cooling is the normalized difference between the temperature in the room and the external temperature. Second, the AC accepts the highest priority requests until the available capacity is consumed; the other requests are deferred. Finally, it sends a signal back to each load either to run (if accepted) or to stand by (if deferred).

Note that the use of suitable priority values eliminates the possibility that a request be deferred indefinitely, so long as there is sufficient capacity to accommodate all the requests within the planning horizon.

3.1.2 Load balancer

The LB spreads the deferred load requests over a given time horizon so as to respect the capacity limits set by the DRM and the constraints of the requests. This is done using a mixed-integer linear optimization problem that minimizes the overall operational cost subject to the capacity and operational constraints, and making sure that each load is scheduled over a proper number of consecutive time steps within the scheduling horizon.

Consider a problem consisting of n loads to be scheduled in a horizon containing m time frames. The loads are denoted by $I = \{1, \ldots, n\}$, and the time frames by $J = \{1, \ldots, m\}$. (For simplicity we assume that the time frames are all equal.) We also let P_i denote the power consumption of load i when it operates, and K_j denote the cost of energy per time frame. It follows that $F_{ij} = P_i K_j$ is the cost incurred if load i operates during time frame j.

Next we define the binary variables to represent the scheduling decisions:

$$x_{ij} = \begin{cases} 1 & \text{if load } i \text{ operates during time frame } j \\ 0 & \text{otherwise,} \end{cases}$$

for $i \in I, j \in J$.

We also need to model the fact that loads may operate for longer than one time frame. For this we introduce a second set of binary variables that track the time frame when each load starts operating:

$$d_{ij} = \begin{cases} 1 & \text{if load } i \text{ is scheduled to start operating during time frame } j \\ 0 & \text{otherwise,} \end{cases}$$

for $i \in I, j \in J$. Using these variables, we can enforce that when load i starts, it must be assigned τ_i consecutive time frames to reach completion. This is done with the constraints (1c) below.

We can now state the optimization problem for the LB:

$$\begin{aligned}
\min \quad & \sum_{i,j} F_{ij} x_{ij} && && \text{(1a)}\\
\text{s.t.} \quad & \sum_i P_i x_{ij} \leq C_j, && \forall j \in J && \text{(1b)}\\
& d_{ij} \leq x_{it} && && \\
& \quad t = j, j+1, \cdots j + \tau_i - 1, && \forall j \in J, \ \forall i \in I && \text{(1c)}\\
& \sum_j d_{ij} = 1, && \forall i \in I && \text{(1d)}\\
& x_{ij} = 0, && \forall i \in I, \ \forall j \notin (T_i^{earliest}, T_i^{latest}) && \text{(1e)}\\
& d_{ij} \geq 0, && \forall i \in I, \ \forall j \in J && \text{(1f)}\\
& x_{ij} \in \{0, 1\}, && \forall i \in I, \ \forall j \in J && \text{(1g)}
\end{aligned}$$

where C_j is the power capacity available for time frame j, and $T_i^{earliest}$ and T_i^{latest} are respectively the earliest and latest start times for load i.

In this optimization problem, apart from the non-negativity and binarity of variables, there are four sets of constraints, namely:

(1b) The total power consumption at each time frame has to respect the capacity limit.

(1c) For each load, the required number of contiguous frames is allocated so that it can be completed before the deadline.

(1d) Each load is scheduled only once.

(1e) Each load is scheduled within the allowed operation time interval.

3.1.3 Demand response manager

The DRM tracks quality of service issues such as the rate of rejection of requests. Based on all this information, the DRM can react to price signals from the grid, or make requests to the grid for additional power capacity when necessary. This design allows the DRM to accommodate different pricing strategies, such as critical-peak, time-of-use, or real-time pricing. There are different models for DR management proposed in the literature, e.g. [11, 12, 17, 39]; they can be readily integrated into this architecture.

The Load Forecaster (LF) is an auxiliary module that provides the DRM and the LB with some of the information they need. For example, a reliable load forecast makes it possible to plan the operation of appliances to take advantage of lower price periods in the context of TOU or Multi-TOU pricing, or to make better decisions about capacity booking in a TLOU context. A comprehensive review of forecasting methods from classical time series to more sophisticated machine learning tools can be found in [41]. Jain et al. [23] use support vector regression to evaluate the impact of the time and space granularity inside a multi-family unit. Specifically in buildings with regular patterns of consumption, a short-term forecasting method for aggregated loads is presented in [30]. The importance of an accurate estimation for exploiting the DR potential is highlighted in [32] where an error analysis for different load estimation methods used in real-world operations is presented. We refer the reader to the recent INFORMS Tutorial [43] for a survey of approaches to modeling uncertain loads, and a discussion of the importance of learning-oriented models.

Load estimation methods designed to support the application of TLOU pricing are presented in the next section.

4 Load estimation

There is a significant body of literature on forecasting the variations in consumption. Load estimation methods can be categorized as bottom-up or top-down [42], and within these two categories different approaches have been used to estimate the energy demand. For example, the well-known k-means approach was used in [7] to estimate future load profiles using complete and incomplete past information, and the use of Logistic and Poisson regression in [40] to estimate energy demand in a large aggregated population.

We describe here approaches to estimate power capacity profiles for a building with the perspective of supporting the implementation of a TLOU tariff as described in Section 2.4 and the use of a building EMS (Section 3). From a DR perspective, the objective is to obtain accurate estimates of the power capacity needed at each time frame to meet the needs of the building for the next planning horizon (e.g., the next day), thus supporting the operation of the EMS, and hence the provision of DR via the DRM.

The idea is that a capacity profile establishes the basis for a tradeoff between the building's energy requirements and the grid's operational decisions. This approach is consistent with the framework proposed in [28] in which a DR aggregator defines a capacity constraint and the consumer consequently minimizes its costs. A related framework is presented in Section 6 below. By setting in advance a level of consumption for each time frame in the planning horizon under a TLOU tariff, the consumer saves money as long as it succeeds in scheduling the building's loads to keep the total power consumption below the set level. From the grid's perspective, the knowledge of this capacity profile for a large number of buildings helps plan the power system operations more efficiently without needing to know any information about the individual loads. Moreover if a building exceeds the agreed capacity, the higher price provides a compensation for the actions that the grid may have to take to accommodate this change in demand.

Because a building typically generates a set of heterogenous loads, it is essential to classify them in terms of both their level of power consumption and their frequency of operation. Following previous work in the literature, see e.g. [12, 17, 25], we use a classification of loads into three categories as follows:

- **Non-flexible activity-based loads**, or simply non-flexible loads, must be served immediately at any time to keep them operating or on standby. This includes lighting, cooking, and electronic devices. Non-flexible loads cannot be deferred and thus must be taken into account when computing the (remaining) available capacity to satisfy other demands at every point in time.
- **Thermal loads** arise from devices ensuring that specific temperatures are maintained. This includes heating, air conditioning, refrigerators, and water heaters. Unlike non-flexible load however, their power consumption can be modulated (within given limits) and in some cases interrupted. Hence their operation can be managed.
- **Flexible Activity-based loads**, or simply activity-based loads, typically have a fixed duration and power consumption, but their number and frequency directly depends on the decisions of the current building consumers. Examples of such devices include washing machines, dryers, and dishwashers.

Although they can represent an important amount of energy consumption, most non-flexible loads consume power levels that are low in comparison to those of the other two categories. We therefore focus on the two other categories that impact power consumption more significantly. We describe here approaches to estimate power capacity profiles for thermal and activity-based loads in a building. These approaches exploit the structure of the estimation problem for these two types of loads. Once this is done, we can deduce a total capacity profile that accounts for all three types of loads, and hence for the total load of the building for the next planning horizon.

4.1 Profile estimation for thermal loads

We describe in this section a recently proposed approach to estimate the power capacity profile required to meet the thermal load consumption of a building [21]. This approach is based on a classification strategy, which means that a capacity level for each time frame is selected from a prespecified set of possible levels. Some related energy problems have been treated in this way in the literature, for example price forecasting in [46] and wind power ramp events in [45].

To estimate the power level required to meet the expectations of the people in the building, we use a measure of *quality of service (QoS)*. This metric is especially important in a context of TOU or TLOU pricing because the consumer can profit from the cheaper time frames by reshaping the load curve while ensuring the desired QoS. We define the QoS for each time frame j as follows:

$$QoS_j = \begin{cases} \dfrac{\sum\limits_{i=1}^{I}\sum\limits_{t=1}^{S} x_{i,t}}{N_j} \times 100\% & N_j > 0 \\ 100\% & N_j = 0, \end{cases} \tag{2}$$

where

$$x_{i,t} = \begin{cases} 1 & \text{if a request from load } i \text{ is accepted in time step } t \\ 0 & \text{otherwise,} \end{cases}$$

$$r_{i,t} = \begin{cases} 1 & \text{if a request is created by load } i \text{ in time step } t \\ 0 & \text{otherwise,} \end{cases}$$

and $N_j = \sum_{i=1}^{I}\sum_{t=1}^{S} r_{i,t}$.

The QoS consolidates, at the scale of a time frame, the performance of the operational system over a number of time steps. We also assume that the capacity level used by the AC of the EMS for a given time frame, is to be chosen from a discrete set Ω of prespecified levels of capacity that are compatible with the loads. This is realistic because thermal loads such as air-conditioning and heating units have a constant level of power consumption when they operate [26], and we do not want to operate fractions of loads.

Our objective is now determine, for a given time frame j, the smallest capacity level in Ω that suffices to operate the thermal loads in the building so that, given the external temperature, the prescribed QoS can be met. We address this question using the following classification problem:

$$\Phi(T_j^e, QoS_j) = C_j \tag{3}$$

that takes as inputs the external temperature T_j^e and the user-selected QoS_j, and returns the suitable choice of $C_j \in \Omega$.

The classification problem (3) is solved in [21] using a three-step approach:

1. Selection of the training set from historical data;
2. Function fitting;
3. Final classification.

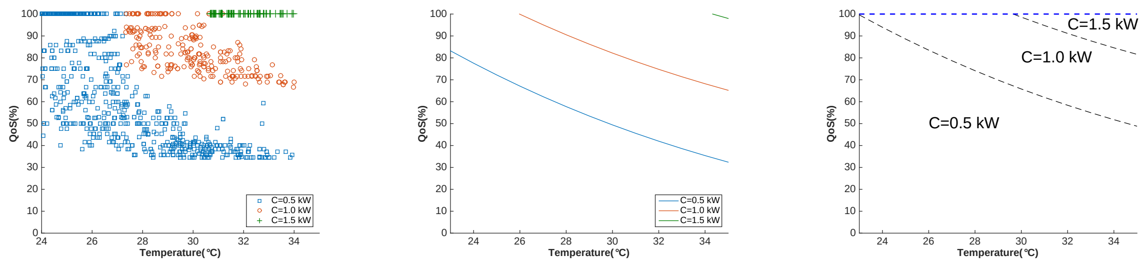
We sketch here Steps 2 and 3, and refer the reader to [21] for the full approach.

Let us first consider Step 2. Suppose that we have the set of data depicted on the graph of QoS versus external temperature in Figure 2 for three different capacity levels. The idea is to fit a sigmoid function of the form:

$$\widehat{QoS}_j = \frac{\beta_1}{1 + e^{\beta_2 T_j^e}} + \beta_3 C_j + \beta_4, \tag{4}$$

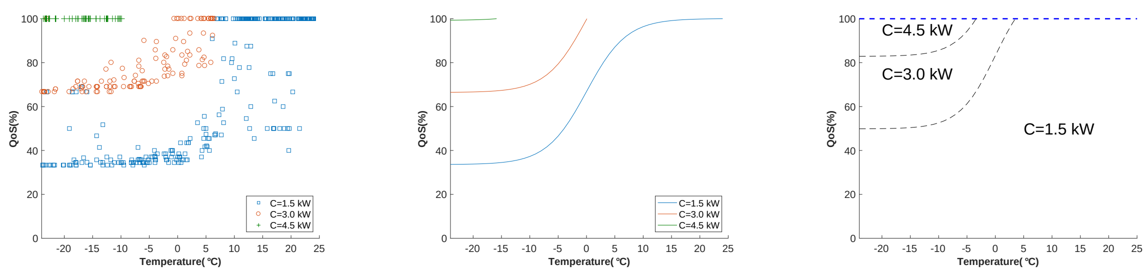
where $\widehat{QoS}_j$ is the quality of service at time frame j as computed by the prediction model.

A sigmoid function helps to represent the asymptotic extremes and monotonic behavior of the QoS. Using a linear function would capture the monotonic condition but not the asymptotic extremes. For cooling systems one uses a monotonically decreasing sigmoid function over the interval of external temperature where cooling is required, as shown in Figure 2, whereas an increasing sigmoid is used for heating systems, as in Figure 3.



QoS vs external temperature — Sigmoid fitting — Resulting classification

Figure 2: Example for cooling



QoS vs external temperature — Sigmoid fitting — Resulting classification

Figure 3: Example for heating

Once a sigmoid function is fitted for each capacity level, it approximates the relationship between that capacity, the external temperature and the QoS. Because the QoS and the temperature are continuous values while the capacity belongs to a discrete set, the classification is done by identifying the areas bounded by the resulting average of two contiguous sigmoid curves. This is based on the idea of nearest neighbor classifier [8].

4.2 Profile estimation for activity-based loads

We now turn our attention to the power capacity required for the activity-based loads. These loads have to be treated differently because of the stochastic nature of the users' behaviour, and hence of their scheduling. However, using a stochastic optimization model requires sufficient knowledge about the loads to be effective in practice. This difficulty is addressed in [20] by proposing a heuristic approach that uses historical data to build an approximate load profile. This load profile can be refined as more data becomes available, and the information acquired through this approach can subsequently inform the use of a stochastic optimization model to estimate the desired capacity profile.

4.2.1 Heuristic to approximate activity-based load profiles

The input to the heuristic is a matrix $\mathcal{H} \in \mathbb{R}^{\mathcal{N}\times|\mathcal{T}|}$ of historical information about the power consumption in the previous N days, where $|J|$ is the number of time frames per day. As a small example, suppose that $N = 3$, $|J| = 12$, and

$$\mathcal{H} = \begin{pmatrix} 0 & 0 & \mathrm{h}_1 & 0 & 0 & 0 & 0 & 0 & \mathrm{h}_5 & 0 & 0 & 0 \\ 0 & \mathrm{h}_2 & 0 & 0 & 0 & \mathrm{h}_4 & 0 & 0 & \mathrm{h}_5 & 0 & 0 & \mathrm{h}_7 \\ 0 & 0 & \mathrm{h}_3 & 0 & 0 & 0 & 0 & 0 & \mathrm{h}_5 & \mathrm{h}_6 & \mathrm{h}_6 & \mathrm{h}_6 \end{pmatrix}$$

The first step of the heuristic is to divide $\mathcal{H}$ into contiguous submatrices by clustering time frames so that each sub-matrix contains only columns with no load (all the entries are zero) or only columns with some load. For our small example, the submatrices are as follows, where the submatrices with load have entries in bold:

$$\mathcal{H} = \begin{pmatrix} 0 & \mathbf{0} & \mathbf{h_1} & 0 & 0 & \mathbf{0} & 0 & 0 & \mathbf{h_5} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{h_2} & \mathbf{0} & 0 & 0 & \mathbf{h_4} & 0 & 0 & \mathbf{h_5} & \mathbf{0} & \mathbf{0} & \mathbf{h_7} \\ 0 & \mathbf{0} & \mathbf{h_3} & 0 & 0 & \mathbf{0} & 0 & 0 & \mathbf{h_5} & \mathbf{h_6} & \mathbf{h_6} & \mathbf{h_6} \end{pmatrix}$$

The second step of the heuristic computes the capacity profile by assigning the average historical consumption from each submatrix to all its time frames. For our example, the resulting capacity profile is:

$$\left[0,\ \frac{\mathrm{h}_2}{3},\ \frac{\mathrm{h}_1+\mathrm{h}_3}{3},\ 0,\ 0,\ \frac{\mathrm{h}_4}{3},\ 0,\ 0,\ \mathrm{h}_5,\ \frac{\mathrm{h}_6}{3},\ \frac{\mathrm{h}_6}{3},\ \frac{\mathrm{h}_6+\mathrm{h}_7}{3}\right].$$

This process of identification of the submatrices is key to the formal proof that the algorithm terminates. A detailed description of this algorithm is given in [20] where the finite termination of this procedure is also proved.

The choice of N can have an important impact in the capacity profile determined by the heuristic. If N is too small then the result will have (very) limited usefulness because it is based on insufficient information. If N is too large then rare events that not represent the user behavior are more likely to affect the outcome. Computationally one can observe that as N increases, the number of submatrices tends to stabilize. A possible practical approach is therefore to include the columns for more days until there is essentially no change in the resulting capacity profile.

4.2.2 Activity-based load profiles using stochastic optimization

Under certain assumptions, we can express the capacity profile estimation problem for activity-based loads as a two-stage stochastic optimization problem. In particular, this formulation assumes a TLOU tariff.

We suppose that for each device m, the duration of operation L_m and the power consumption level P_m are fixed, and that the starting time frame follows a normal distribution discretized over the $|J|$ time frames of the planning horizon. Let $Pr(X_{mj} = 1)$ denote the probability that device m starts in time frame j, then the probability that device m is operating in time frame j equals

$$Pr(\tilde{X}_{mj} = 1) = \sum_{j-L_m}^{j} Pr(X_{mj} = 1).$$

The possible accumulation of more than one load operating during the same time frame results in many different possible scenarios. For a given scenario i consisting of a set of loads operating at the same time, the probability that i occurs at time frame j equals

$$Pr_{i,j} = \prod_{m\in i} Pr(\tilde{X}_{mj} = 1) \prod_{m\notin i} (1 - Pr(\tilde{X}_{mj} = 1)).$$

In practice many of these probabilities will be almost zero. We can remove from consideration the combinations (i, j) for which the probability is below some threshold determined by the user.

The two-stage stochastic optimization formulation works as follows:

- the first stage sets the capacity profile, i.e., the power capacity required for each time frame, and
- the second stage meets demand at minimum cost, subject to the capacity profile set.

The notation used is as follows. The sets A and Q represent the intervals of the step functions previously introduced in Figure 1. The variable c_{aj} corresponds to the capacity booked in the first stage in each time frame j, while x^L_{iaj} and x^H_{iqj} account for the consumption below and above said capacity for each scenario i in the second stage. The binary variables ϕ_{aj} and δ_{qj} are used in identifying the inverval where the capacity belongs. Finally, the costs K^L_{aj}, K^H_{qj} and K^F_j and are those of the TLOU tariff from Section 2.4.

The objective function (5) is the sum of the booking cost, the expected cost of consumption within the capacity profile (lower tariff), and the expected cost of consumption above the levels of the profile (higher tariff):

$$\sum_{j\in J}\sum_{a\in A} K^F_j c_{aj} + \sum_{j\in J}\sum_{a\in A}\sum_{i\in I(j)} Pr_{ij} K^L_{aj} x^L_{iaj} + \sum_{j\in J}\sum_{q\in Q}\sum_{i\in I(j)} Pr_{ij} K^H_{qj} x^H_{iqj} \tag{5}$$

Constraints (6) and (7) ensure that the chosen capacity lies on the step functions defining the tariffs (as illustrated in Figure 1):

$$\sum_{a\in A} \phi_{aj} = 1 \quad \forall j \in J \tag{6}$$

$$\sum_{q\in Q} \delta_{qj} = 1 \quad \forall j \in J \tag{7}$$

Constraints (8) and (9) set the lower and upper bounds for each interval of the step functions. Here we introduce the auxiliary variable $\bar{c}_{qj}$ for the capacity in the higher tariff step cost function.

$$\phi_{aj} C^L_{aj} \le c_{aj} \le \phi_{aj} C^L_{a+1j} \quad \forall a \in A \mid a <\mid A \mid -1, j \in J \tag{8}$$

$$\delta_{qj} C^H_{qj} \le \bar{c}_{qj} \le \delta_{qj} C^H_{q+1j} \quad \forall q \in Q \mid q <\mid Q \mid -1, j \in J \tag{9}$$

Constraint (10) establishes the relationship between the capacity and the auxiliary variable.

$$\sum_{a\in A} c_{aj} - \sum_{q\in Q} \bar{c}_{qj} = 0 \quad \forall j \in J \tag{10}$$

Constraints (11) and (12) ensure the lower level consumption and the demand satisfaction for each scenario respectively.

$$x^L_{iaj} \le c_{aj} \quad \forall i \in I(t), a \in A, j \in J \tag{11}$$

$$\sum_{a\in A} x^L_{iaj} + \sum_{q\in Q} x^H_{iqj} \ge D_{ij} \quad \forall i \in I(t), j \in J \tag{12}$$

Finally constraints (13) and (14) enforce the nature of each of the variables (non-negative or binary):

$$x^L_{iaj}, x^H_{iqj}, c_{aj}, \bar{c}_{qj} \ge 0, \quad \forall i \in I(j), a \in A, q \in Q, j \in J \tag{13}$$

$$\phi_{aj}, \delta_{qj} \in \{0, 1\} \quad \forall a \in A, q \in Q, j \in J \tag{14}$$

The stochastic optimization formulation is thus:

$$\min \quad (5) \quad \text{subject to} \quad (6) - (14) \tag{15}$$

As it stands, the above formulation allows the capacity profile to have a different level at each time frame. A more realistic requirement is that the capacity profiles should be constant over certain time windows, for example the groups of time frames corresponding to the same price in the TLOU/TOU tariff. This requirement can be enforced for a subset $\tau^\omega \subset J$ using constraints (16):

$$c_{aj} = c_{aj'} \quad \forall a \in A, \quad j, j' \in \tau^\omega \mid j \neq j', \quad \omega \in \Omega. \tag{16}$$

Models (5)–(14) and (5)–(16) consider the aggregated consumption scenarios that are computed from the individual load distributions. Although this aggregation can result in a large number of scenarios (even after eliminating near-zero probabilities), the problem can be solved efficiently because the binary variables only represent the interval where the chosen capacity belongs in order to identify the corresponding tariffs.

5 Local generation, energy storage and electric vehicles

Technologies for local power generation, particularly solar panels and wind turbines, have progressed dramatically in recent years, and are being adopted by many sectors of the economy, including residential consumers. At the same time, rapid improvements in the size and efficiency of energy storage technologies, particularly with respect to batteries, are already having major impacts in the transportation sector, and are poised to significantly impact the residential, commercial, and institutional sectors in the near future. Two manifestations of this impact are the increasing numbers of electric vehicles, and the design of batteries for use within a building.

We present here the optimization model proposed in [31] that considers a residential building and integrates EVs with the objective of reducing purchases of energy from the grid. Specifically the objective is to schedule the consumption loads in the building as well as the charging of the EVs so as to miminize the amount of power procured from the grid. The only batteries considered are those of the EVs, and for simplicity we omit the EV battery ramping constraints. An example on the use of dedicated batteries in the operation of a building is presented in the next Section.

5.1 Optimization model

Let $Grid_t$ denote the amount of energy to be procured from the grid at time frame t. The objective function is then to minimize $\sum_t Grid_t$ over the planning horizon (typically one day) subject to a number of constraints:

Next we let $x_{i,t}$ be the variable representing the amount of power consumed by load i during time frame t. For every non-flexible load i, $x_{i,t}$ must equal the required power. This can be viewed as a first set of constraints, or equivalently the variables $x_{i,t}$ for non-flexible loads can be eliminated by replacing them with the corresponding power level.

For thermal loads i, we have some flexibility: the operating power level has to be within prescribed limits at each time frame, and the total amount of power (or equivalently energy) has to be provided during the planning horizon:

$$\underline{x_{i,t}} \leq x_{i,t} \leq \overline{x_{i,t}}, \quad \sum_t x_{i,t} = Pow_i \tag{17}$$

where $\underline{x_{i,t}}$ and $\overline{x_{i,t}}$ are the bounds (which may vary from one time frame to another).

An interesting feature of this model is the way it handles the scheduling of the activity-based loads. The assumption here is that while the starting time of each of these can be scheduled, their power consumption pattern is fixed once it started operating. Under this assumption, the idea is to represent the possible scheduling choices for each of these loads using a matrix that has one column embedding this pattern within the planning horizon according to every allowed starting time frame for the load. Specifically, suppose that load i has to operate for τ_i time frames and that the power required during each of those time frames is given by $h_1, h_2, \ldots, h_{\tau_i}$. If load i can start at any time frame and there are 24 time frames in the planning horizon, then the matrix will have the form:

$$\mathcal{S}_\rangle = \begin{pmatrix} h_1 & 0 & 0 & \dots & h_2 \\ h_2 & h_1 & 0 & \dots & h_3 \\ h_3 & h_2 & h_1 & \dots & h_4 \\ \vdots & \vdots & \vdots & & \vdots \\ h_{\tau_i-1} & h_{\tau_i-2} & h_{\tau_i-3} & \dots & h_{\tau_i} \\ h_{\tau_i} & h_{\tau_i-1} & h_{\tau_i-2} & \dots & 0 \\ 0 & h_{\tau_i} & h_{\tau_i-1} & \dots & 0 \\ 0 & 0 & h_{\tau_i} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & h_1 \end{pmatrix}$$

Note that this matrix has a circulant structure, i.e., each row is equal to the previous row rotated one element to the right.

With this setup, scheduling load i is equivalent to choosing one of the columns of the matrix. This is done using a vector s_i of binary variables of length equal to the number of columns in $\mathcal{S}_\rangle$ and such that exactly one of the variables in s_i is equal to 1; then $\mathcal{S}_\rangle \int_\rangle$ is the column vector with the schedule for load i. Therefore the constraints for activity-based load i are:

$$x_{i,t} = (\mathcal{S}_\rangle \int_\rangle)_\sqcup, \quad \sum_\sqcup \int_{\rangle,\sqcup} = \infty. \tag{18}$$

The next set of constraints ensures the equilibirum between power provision and demand:

$$Gen_t - \sum_i x_{it} + G_t - \sum_v \Phi_{v,t} = 0, \quad \forall t \tag{19}$$

where Gen_t is the amount of power coming from the solar panels, or more generally from local generation, and $\Phi_{v,t}$ is the amount of power dedicated to charging the battery of EV v during time frame t.

One more set of constraints enforces energy balance for each of the batteries:

$$SOC_{v,t} = SOC_{v,t-1} + \Phi_{v,t} - \phi_{v,t} = 0 \tag{20}$$

The last constraints enforce the appropriate bounds or binarity of the variables:

$$s_{i,t} \in \{0.1\}, \tag{21}$$

and

$$0 \leq Grid_t, \quad 0 \leq \Phi_{v,t} \leq \overline{\Phi_{v,t}}, \quad 0 \leq SOC_{v,t} \leq \overline{SOC_v} \tag{22}$$

where $\overline{SOC_v}$ is the energy capacity of the battery of EV v.

As can be seen from (20) above, the model as presented in [31] allows for the EV batteries to both charge and discharge during the same time frame. This may be undesirable, depending on the length of the time frame. One way to prevent it from happening is to force (at least) one of them to equal zero using a binary variable $EV_{v,t}$ and the constraints:

$$\Phi_{v,t} \leq \overline{\Phi_{v,t}} EV_{v,t}, \quad \phi_{v,t} \leq (1 - EV_{v,t})\Phi_{v,t-1} \tag{23}$$

where $\overline{\Phi_{v,t}}$ is the charging availability of EV v at time frame t.

We refer the reader to [31] for additional details and experimental results.

6 Building-level aggregation

While the residential and commercial sectors represent a major part of the total consumer demand, the ability for a single household to provide DR is naturally limited. It is therefore important to study how to aggregate their DR potential so that it can be more effective in practice. Moreover, an entity capable of coordinating the DR actions of these consumers could ensure that the rebound peaks mentioned in Section 2.3 do not happen.

Aggregation in the field of energy is a well-researched topic. Much previous work is on the aggregation of demand, possibly in the presence of storage and/or distributed generation, has been done in the context of microgrids. The comprehensive review presented in [37] introduces various aspects of microgrids, including operation, investment, generation technologies, communications requirements, and grid-support and islanding capabilities.

We present here the framework proposed in [19] to facilitate the coordination process in a smart building with multiple units that manage their energy consumption independently, for example using an EMS such as the one described in Section 3. This includes the case of a building with multiple households. The framework provides DR by aiming to keep its energy consumption within a given capacity level, and is thus particularly suited to take advantage of a TLOU tariff (see Section 2.4). The framework also accounts for the smart building having a shared battery (or other capacity for energy storage) and a set of solar panels. Note that the battery and solar panels are managed by the building, and they are resources that are shared among all the housing units. This increases the potential for the building, and thus for the units within it, to profit from the opportunities to provide DR.

One important assumption in this framework is that the units have the choice to participate or not in the provision of DR. The idea is that each unit submits preferences indicating when and by how much it is prepared to shift or reduce consumption. Using this information, and based on the tariff (or other incentive program) in effect, the framework uses an optimization model to decide if the building should respond to a DR request. If responding, the building commits to adjusting consumption to a lower capacity level through using the battery and/or allocating capacity reductions to the building's units that agree to participate. Each of these units then plans its energy consumption accordingly, and is allocated a proportion of the financial benefit corresponding to its share of the DR provided.

To account for the user preferences, the framework optimizes two objectives, namely to minimize the shifted load, and to minimize the total cost for the provision of electricity to the units in the building. The shifted load is as a measure of the level of dissatisfaction perceived by the units, and thus these objectives inevitably conflict in the sense that there is usually no solution that optimizes them simultaneously.

Other approaches in the literature have accounted more than one objective, often by minimizing the total operational cost while including user preferences via constraints and/or costs that approximate the level of satisfaction. This way of dealing with conflicting objectives is one among several techniques in multiobjective optimization [16]. All of them trade-off between the conflicting objectives.

The framework applies the approach of *compromise programming*. This approach makes use of the concept of utopia point, which is a point where all the objectives achieve their individual optima. Since the objectives conflict, the utopia point is infeasible. The motivation for compromise programming is to find a point on the Pareto front that is a good approximation of the utopia point, and one way to do his is by minimizing the Euclidean distance with respect to the utopia point [29]. The resulting nearest point to the utopia point is the trade-off between the two objectives chosen by the framework.

6.1 Smart building planning module

The framework is implemented within a *smart building planning module* for the smart building. This module receives day-ahead information from the building's units and resources: user preferences, forecasts of energy demand and solar radiation, battery state of charge, and scheduled DR requests from the grid.

Once all the information has been gathered, the planning model finds the compromise solution between shifted load and total cost. First, it solves the optimization problem for each of the individual objectives to compute the utopia point. Second, it solves a mixed binary quadratic optimization problem to find the closest feasible point to the utopia point. Both optimization problems can be solved efficiently using off-the-shelf solvers, either commercial or open source. The solution of the last optimization problem provides individual plans that specify for each time frame the amount of energy to be drawn from the grid and the solar panels, the use of the battery, and the shifted load for each unit. This is similar to capacitated lot sizing with backlogging [38], where the objective is to minimize the sum of the production, storage, and backlogging costs.

We present next the details of the mixed binary quadratic optimization problem. For simplicity, we omit the aspects of the model accounting for the cycles in the operation of the battery. We refer the reader to [19] for additional details, including the analogy with capacitated lot sizing, and detailed experimental results.

6.2 Optimization model

6.2.1 Objective functions

There are three objective functions to consider.

The first one is the total load shifted during the planning horizon:

$$f_{TLS} = \sum_{j \in J} \sum_{t \in T} y_{jt} \tag{24}$$

where y_{jt} is the accumulated unmet demand at the end of period t for unit j.

The second objective function is the total cost, consisting of the cost of energy bought from the grid, the cost of using the battery and solar panels, and the incentive paid for the provision of DR:

$$f_{TC} = \sum_{i \in I} \sum_{j \in J} \sum_{t \in T} K_{it} x_{ijt} + \sum_{j \in J} \sum_{t \in T} B s^{+}_{jt} + \sum_{j \in J} \sum_{t \in T} F g_{jt} - \sum_{j \in J} \sum_{t \in T} L_t r_{jt} \tag{25}$$

where the quantities involved in each term are as follows, for energy level i, time frame t, and unit j:

- K_{it} is the cost of one unit of energy bought from the grid, and x_{ijt} is the amount of energy bought from the grid
- B is the cost per energy unit of charging the battery, and s^{+}_{jt} is the amount of energy injected into it
- F is the cost per energy unit obtained from the solar panels, and g_{jt} is the amount of energy produced by them
- L_t is the amount paid to the units for providing one unit of DR, and r_{jt} is the amount of DR provided.

The third objective is the Euclidean distance from any point to the utopia point. Let $\tilde{u}_{TLS}$ and $\tilde{u}_{TC}$ denote the optimal solutions obtained for each of the previous two objective functions under the constraints described below. These two values give the utopia point.

Because we are dealing with objectives with different units and different orders of magnitude, it is necessary to normalize their values. For this purpose we need to also compute the nadir point $(\hat{u}_1, \hat{u}_2)$. The nadir point corresponds to the worst value of each objective function along the Pareto-optimal front. Note that this is in general different from the point representing the worst objective values of the entire search space. We refer the reader to [16] for more information on multiobjective optimization.

The nadir and utopia points provide tight upper and lower bounds on the values of the first two objectives, and are used to scale the distance in the third objective as follows:

$$\min \quad \left(\frac{f_{TLS} - \tilde{u}_{TLS}}{\hat{u}_{TLS} - \tilde{u}_{TLS}} \right)^2 + \left(\frac{f_{TC} - \tilde{u}_{TC}}{\hat{u}_{TC} - \tilde{u}_{TC}} \right)^2 \tag{26}$$

6.2.2 Constraints

For all the constraints, we use i for the capacity level, t for the time frame t, and j for the unit.

Constraints (27) account for the shifting preferences of each unit, i.e., they enforce the maximum accumulated shifted load for unit j throughout the planning horizon:

$$\sum_{t \in T} y_{jt} \leq Y_j \qquad \forall \ \ j \in J \tag{27}$$

where Y_j is the backlogged demand acceptable. Each unit can also specify, via (28), the maximum acceptable unmet demand at the end of the horizon:

$$y_{jn} \leq \hat{Y}_j \qquad \forall \ \ j \in J \tag{28}$$

where $\hat{Y}_j$ is the maximum backlogged demand at the end of the planning horizon. In other words, the model allows a unit to be flexible about when its demand is satisfied but strict about having it met by the end of the day.

The amounts of energy drawn from the grid are accounted for by constraints (29) and (30):

$$x_{1jt} \leq C^L \qquad \forall \ \ j \in J, \ \ \forall \ \ t \in T \tag{29}$$

$$x_{2jt} \leq C_j^H \qquad \forall \ \ j \in J, \ \ \forall \ \ t \in T \tag{30}$$

where C^L is the available capacity in the lower level of the TLOU tariff, and C_j^H is the maximum capacity in the higher level (or a large number, as applicable).

Constraints (31) tracks the state of charge of the battery on the basis of each unit in the building. In other words, soc_{jt} is the amount of energy in the shared battery available to unit j at time frame t:

$$soc_{jt} = soc_{jt-1} + s_{jt}^+ - s_{jt}^- \quad \forall \, j \in J, \, \forall \, t \in T \tag{31}$$

where s_{jt}^- is the energy extracted from the battery.

Constraints (32) and (33) model the capacities of the solar panels and the battery respectively:

$$\sum_{j \in J} g_{jt} \leq G_t^{max} \qquad \forall \ \ t \in T \tag{32}$$

$$\sum_{j \in J} soc_{jt} \leq S^{max} \qquad \forall \ \ t \in T \tag{33}$$

where G_t^{max} is the energy available from solar panels, and S^{max} is the capacity of the battery.

Constraints (34) ensure the energy balance of the whole system:

$$\sum_{i \in I} x_{ijt} + g_{jt} + y_{jt} + \Gamma s_{jt}^- = y_{jt-1} + D_{jt} + s_{jt}^+ \quad \forall \ \ j \in J, \ \ \forall \ \ t \in T \tag{34}$$

where Γ is the efficiency of the battery, and D_{jt} is the demand of unit j (prior to the provision of DR).

Constraints (35) describe the physical energy flow toward the battery. They are necessary to avoid charging the battery with backlogged load:

$$g_{jt} + \sum_{i \in I} x_{ijt} - s_{jt}^+ \geq 0 \qquad \forall \ \ j \in J, \, \forall \ \ t \in T \tag{35}$$

Constraints (36) and (37) ensure a proportional use of the shared resources with respect to the total demand of each unit:

$$\Psi_{sol}^{min} \sum_{t \in T} D_{jt} \leq \sum_{t \in T} g_{jt} \leq \Psi_{sol}^{max} \sum_{t \in T} D_{jt} \quad \forall \ \ j \in J \tag{36}$$

$$\Psi_{bat}^{min} \sum_{t \in T} D_{jt} \leq \sum_{t \in T} soc_{jt} \leq \Psi_{bat}^{max} \sum_{t \in T} D_{jt} \quad \forall \;\; j \in J \tag{37}$$

They are needed to ensure a fair allocation of the shared resources. The choice of the parameters Ψ_{sol}^{min}, Ψ_{sol}^{max}, Ψ_{bat}^{min}, and Ψ_{bat}^{max} requires some care. We refer the reader to [19] for a detailed discussion.

Constraints (38) and (39) model the building's response to the DR requests. If the building agrees to provide DR, each participating unit j contributes r_{jt} to the grid's load reduction requirement:

$$\sum_{j \in J} r_{jt} = DR_t \phi_t \quad \forall \; t \in T \tag{38}$$

where DR_t is the load decrease requested by the grid, and ϕ_t is a binary variable indicating if the building agrees to provide DR in time frame t ($\phi_t = 1$) or not ($\phi_t = 0$).

Furthermore, constraints (39) allow the units to reduce their consumption below C^L. If $D_{jt} - r_{jt} \leq C^L$ then the consumption will stay within the capacity available at a lower price of the TLOU tariff. If $C_j^H + C^L \geq D_{jt} - r_{jt} \geq C^L$ then the user will consume C^L units at the lower price and the additional energy at the higher price.

$$x_{1jt} + x_{2jt} \leq (C^L + C_j^H)(1 - \phi_t) + D_{jt}\phi_t - r_{jt} \quad \forall \; j \in J, \forall \; t \in T \tag{39}$$

Finally, constraints (40) and (41) are the nonnegativity and binarity constraints:

$$x, y, soc, s^+, s^-, g, r, \lambda \geq 0 \tag{40}$$

$$\alpha_t, z_t, \phi_t \in \{0, 1\} \tag{41}$$

The combination of the individual decisions to accept to shift load contributes to the peak reduction and to the building's DR capabilities. Even in cases where the population is not very willing to shift load, the building manages to provide DR thanks to the use of the battery and the electricity obtained from the solar panels. Since in this case the TLOU tariffs and corresponding capacity levels are defined by the utility or the system operator, the proper selection of these parameters avoids rebound peaks.

The full model used in [19] included constraints to bound the number of daily (charging and discharging) cycles allowed in operating the batter. Nevertheless the experimental results showed that the demand profiles and the tariff structure end up conditioning the battery charging and discharging patterns. In other words, once the bound on number of cycles is removed, the actual number of cycles remains low due to the intrinsic number of periods in which it makes sense to charge or discharge the battery.

7 Challenges for the future

The potential of DR as a means to increase the flexibility of the smart grid could bring a great number of benefits. This tutorial illustrated some of the recent OR approaches that could help realize its potential in residential, commercial and institutional buildings. However, there are major challenges to overcome before DR can be fully integrated in the operations of a power grid. We conclude by pointing out three aspects which could benefit from the use of OR techniques:

- There is a *multiplicity of opportunities for DR* to contribute to power system operations. The near-totality of the OR literature on DR is concerned with demand peak reduction. While this is an important concern, DR is increasingly expected to play a role in the provision of ancillary services, such as voltage control, operating reserves, and phase balancing. For example, an approach to manage residential power consumption taking into account the constraints of the distribution system was proposed in [33]. There is a need for OR models to integrate the variety of potential contributions of DR and contribute to their realization.

- Related to the first aspect above, there is a need for the *economic arrangements in the provision of DR* to better reflect its value in the electricity markets and/or in the system operation. This is particularly a challenge in the residential sector where few customers pay on a RTP basis, and therefore tariffs are only loosely linked to actual market or system conditions. The development of techniques capable of coordinating the DR actions of large numbers of smaller consumers, see e.g. [18, 24], will make it possible to take advantage of economies of scale. Moreover, optimal pricing of DR can be addressed using OR techniques, see e.g. [6].
- The *interactions between DR and the flows in the power grid*, both at the distribution as well as the transmission levels, are mostly not accounted for in OR models for DR. The importance of the non-linearities in Kirchhoff's laws is well known, and have recently been extensively studied in the OR literature for the computation of power flows, see e.g. [13]. Their impact in electricity markets has also been studied, see e.g. [9]. OR models in the area of DR will increasingly have to account for the networks impacts to remain relevant in practice.

In conclusion, building-level DR has the potential to become an important source of flexibility in the smart grid, but a number of technical and economic challenges need to be overcome to fully realize this potential. OR, and optimization in particular, have an important role to play in supporting the development and realization of this potential in the coming years.

References

[1] Survey of Household Energy Use. Technical report, Natural Resources Canada, 2011.

[2] Households and the Environment: Energy Use. Technical report, Statistics Canada, 2013.

[3] 2014 Yearbook of Electricity Distributors. Technical report, Ontario Energy Board, 2015.

[4] Electric Power Annual 2015. Technical report, U.S Energy Information Administration, 2016.

[5] Assessment of Demand Response and Advanced Metering. Technical report, Federal Energy Regulatory Commission, Dic 2012.

[6] Sezin Afsar, Luce Brotcorne, Patrice Marcotte, and Gilles Savard. Achieving an optimal trade-off between revenue and energy peak within a smart grid environment. Renewable Energy, 91:293–301, 2016.

[7] Ali Al-Wakeel, Jianzhong Wu, and Nick Jenkins. K-means based load estimation of domestic smart meter measurements. Applied Energy, 194:333–342, 2016.

[8] Mohamed Aly. Survey on multiclass classification methods. Technical report, California Institute of Technology, pages 1–9, 2005.

[9] Guillermo Bautista, Miguel F Anjos, and Anthony Vannelli. Formulation of oligopolistic competition in ac power networks: An nlp approach. IEEE Transactions on Power Systems, 22(1):105–115, 2007.

[10] Marc Beaudin and Hamidreza Zareipour. Home energy management systems: A review of modelling and complexity. Renewable and Sustainable Energy Reviews, 45:318–335, 2015.

[11] S. Caron and G. Kesidis. Incentive-based energy consumption scheduling algorithms for the Smart Grid. Technical Report CSE-10-003, Dept. of Computer Science and Engineering, Penn State, May 2009.

[12] Lijun Chen, Na Li, Libin Jiang, and Steven H. Low. Optimal demand response: Problem formulation and deterministic case. In Aranya Chakarabortty and Marija Ilic, editors, Control and Optimization Theory for Electric Smart Grids. Springer-Verlag, New York, 2011.

[13] Carleton Coffrin and Pascal Van Hentenryck. A linear-programming approximation of ac power flows. INFORMS Journal on Computing, 26(4):718–734, 2014.

[14] Giuseppe Tommaso Costanzo, Guchuan Zhu, Miguel F. Anjos, and G. Savard. A system architecture for autonomous demand side load management in smart buildings. IEEE Transactions on Smart Grid, 3(4):2157–2165, 2012.

[15] Elisabeth Dütschke and Alexandra-Gwyn Paetz. Dynamic electricity pricing – which programs do consumers prefer? Energy Policy, 59:226–234, 2013.

[16] Matthias Ehrgott. Multicriteria Optimization. Springer Science & Business Media, 2006.

[17] N. Gatsis and G.B. Giannakis. Cooperative multi-residence demand response scheduling. In Information Sciences and Systems (CISS), 2011 45th Annual Conference on, pages 1–6, march 2011.

[18] F. Gilbert, M.F. Anjos, P. Marcotte, and G.Savard. Optimal design of bilateral contracts for energy procurement. European Journal of Operational Research, 246(2):641 – 650, 2015.

[19] J. A. Gómez and M. F Anjos. Collaborative demand-response planner for smart buildings. Technical Report G-2017-15, GERAD, March 2017.

[20] J. A. Gómez and M. F Anjos. Power capacity profile estimation for activity-based residential loads. Technical Report G-2017-39, GERAD, May 2017.

[21] J. A. Gómez and M. F Anjos. Power capacity profile estimation for building heating and cooling in demand-side management. Applied Energy, 191:492–501, 2017.

[22] Karen Herter. Residential implementation of critical-peak pricing of electricity. Energy Policy, 35(4):2121–2130, 2007.

[23] Rishee K. Jain, Kevin M. Smith, Patricia J. Culligan, and John E. Taylor. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy, v 123:p 168–178, 2014.

[24] Arman C Kizilkale and Roland P Malhamé. Mean field based control of power system dispersed energy storage devices for peak load relief. In Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on, pages 4971–4976. IEEE, 2013.

[25] Jan Kleissl and Yuvraj Agarwal. Cyber-physical energy systems: focus on smart buildings. In Proceedings of the 47th Design Automation Conference, DAC '10, pages 749–754, 2010.

[26] M. Kuzlu, M. Pipattanasomporn, and S. Rahman. Hardware demonstration of a home energy management system for demand response applications. IEEE Transactions on Smart Grid, 3(4):1704–1711, Dec 2012.

[27] Sebastian Lehnhoff, Olav Krause, and Christian Rehtanz. Dezentrales autonomes energiemanagement. At-Automatisierungstechnik Methoden und Anwendungen der Steuerungs-, Regelungs-und Informationstechnik, 59(3):167–179, 2011.

[28] K. Margellos and S. Oren. Capacity controlled demand side management: A stochastic pricing analysis. IEEE Transactions on Power Systems, 31(1):706–717, Jan 2016.

[29] R. T. Marler and J. S. Arora. Survey of multi-objective optimization methods for engineering. Structural and Multidisciplinary Optimization, 26(6):369–395, 2004.

[30] Joaquim Massana, Carles Pous, Lloren Burgas, Joaquim Melendez, and Joan Colomer. Short-term load forecasting in a non-residential building contrasting models and attributes. Energy and Buildings, 92:322–330, 2015.

[31] Petra Mesari and Slavko Krajcar. Home demand side management integrated with electric vehicles and renewable energy sources. Energy and Buildings, 108:1–9, 2015.

[32] S. Mohajeryami, M. Doostan, A. Asadinejad, and P. Schwarz. Error Analysis of Customer Baseline Load (CBL) Calculation Methods for Residential Customers. IEEE Transactions on Industry Applications, PP(99):1–1, 2016.

[33] B. Moradzadeh and K. Tomsovic. Two-stage residential energy management considering network operational constraints. Smart Grid, IEEE Transactions on, 4(4):2339–2346, 2013.

[34] Matteo Muratori and Giorgio Rizzoni. Residential demand response: dynamic energy management and time-varying electricity pricing. IEEE Transactions on Power systems, 31(2):1108–1117, 2016.

[35] Matteo Muratori, Beth-Anne Schuelke-Leech, and Giorgio Rizzoni. Role of residential demand response in modern electricity markets. Renewable and Sustainable Energy Reviews, 33:546–553, 2014.

[36] E Pacey. Smart buildings: people and performance. Royal Academy of Engineering, London, UK, 2013.

[37] S. Parhizi, H. Lotfi, A. Khodaei, and S. Bahramirad. State of the art in research on microgrids: A review. IEEE Access, 3:890–925, 2015.

[38] Yves Pochet and Laurence A Wolsey. Production planning by mixed integer programming. Springer Science & Business Media, 2006.

[39] P. Samadi, A. Mohsenian-Rad, R. Schober, V.W.S. Wong, and J. Jatskevich. Optimal real-time pricing algorithm based on utility maximization for smart grid. In Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on, 415–420, oct. 2010.

[40] R. Subbiah, K. Lum, A. Marathe, and M. Marathe. Activity based energy demand modeling for residential buildings. In Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES, pages 1–6, Feb 2013.

[41] L. Suganthi and Anand A. Samuel. Energy models for demand forecasting: A review. Renewable and Sustainable Energy Reviews, 16(2):1223–1240, 2 2012.

[42] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renewable and Sustainable Energy Reviews, 13(8):1819–1835, 10 2009.

[43] Taylor, J. A., and J. L. Mathieu. Uncertainty in Demand Response – Identification, Estimation, and Learning. The Operations Research Revolution. INFORMS, 2015. 56–70.

[44] World Business Council for Sustainable Development. Transforming the market: Energy efficiency in the buildings. Technical report, WBCSD, June 2009.

[45] H. Zareipour, D. Huang, and W. Rosehart. Wind power ramp events classification and forecasting: A data mining approach. In 2011 IEEE Power and Energy Society General Meeting, pages 1–3, July 211.

[46] H. Zareipour, A. Janjani, H. Leung, A. Motamedi, and A. Schellenberg. Classification of future electricity market prices. IEEE Transactions on Power Systems, 26(1):165–173, Feb 2011.