Pervasive Data Analytics for Sustainable Energy Systems

THÈSE N^{O} 6556 (2015)

PRÉSENTÉE LE 24 AVRIL 2015

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE SYSTÈMES D'INFORMATION RÉPARTIS
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Tri Kurniawan WIJAYA

acceptée sur proposition du jury:

Prof. B. Faltings, président du jury
Prof. K. Aberer, directeur de thèse
Prof. T. B. Pedersen, rapporteur
Dr M. Sinn, rapporteur
Prof. J.-Y. Le Boudec, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

Pervasive Data Analytics for Sustainable Energy Systems

Tri Kurniawan WIJAYA

EPFL - Ecole Polytechnique Fédérale de Lausanne

Abstract

With an ever growing population, global energy demand is predicted to keep increasing. Furthermore, the integration of renewable energy sources into the electricity grid (to reduce carbon emission and humanity's dependency on fossil fuels), complicates efforts to balance supply and demand, since their generation is intermittent and unpredictable. Traditionally, it has always been the supply side that has adapted to follow energy demand, however, in order to have a sustainable energy system for the future, the demand side will have to be better managed to match the available energy supply.

In the first part of this thesis, we focus on understanding customers' energy consumption behavior (demand analytics). While previously, information about customer's energy consumption could be obtained only with coarse granularity (e.g., monthly or bimonthly), nowadays, using advanced metering infrastructure (or smart meters), utility companies are able to retrieve it in near real-time. By leveraging smart meter data, we then develop a versatile customer segmentation framework, track cluster changes over time, and identify key characteristics that define a cluster.

Additionally, although household-level consumption is hard to predict, it can be used to improve aggregate-level forecasting by first segmenting the households into several clusters, forecasting the energy consumption of each cluster, and then aggregating those forecasts. The improvements provided by this strategy depend not only on the number of clusters, but also on the size of the customer base. Furthermore, we develop an approach to model the uncertainty of future demand. In contrast to previous work that used computationally expensive methods, such as simulation, bootstrapping, or ensemble, we construct prediction intervals directly using the time-varying conditional mean and variance of future demand.

While analytics on customer energy data are indeed essential to understanding customer behavior, they could also lead to breaches of privacy, with all the attendant risks. The first part of this thesis closes by exploring symbolic representations of smart meter data which still allow learning algorithms to be performed on top of them, thus providing a trade-off between accurate analytics and the protection of customer privacy.

In the second part of this thesis, we focus on mechanisms for incentivizing changes in customers' energy usage in order to maintain (electricity) grid stability, i.e., Demand Response (DR). We complement previous work in this area (which typically targeted large, industrial customers) by studying the application of DR to residential customers. We first study the influence of DR baselines, i.e., estimates of what customers would have consumed in the absence of a DR event. While the literature to date has focused on baseline accuracy and bias, we go beyond these concepts by explaining how a baseline affects customer participation in a DR event, and how it affects both the customer and company profit. We then discuss a strategy for matching the demand side with the supply side by using a multiunit auction performed by intelligent agents on behalf of customers. The thesis closes by eliciting behavioral incentives from the crowd of customers for promoting and maintaining customer engagement in DR programs.

Keywords: smart grid, data analytics, customer segmentation, smart meter, load forecasting, privacy, demand response, sustainability, dynamic pricing, multi-agent system, human behavior, crowdsourcing

Résumé

La croissance continuelle démographique entraînera une augmentation de la demande mondiale énergétique. De plus, l'intégration de sources d'énergie renouvelable dans le réseau électrique (pour réduire les émissions de carbone et la dépendance de l'humanité aux énergies fossiles) rend difficiles les efforts pour équilibrer l'offre et la demande, du fait de la nature intermittente et imprévisible de leur génération. Traditionnellement, l'offre s'adaptait à la demande énergétique. Cependant, dans le cadre d'un future système énergétique durable, la demande devra être mieux gérée pour être assortie à l'offre énergétique disponible.

Dans la première partie de cette thèse nous nous concentrons sur la compréhension du comportement énergétiques des utilisateurs (analyse de la demande). Alors que des information sur la consommation énergétique des clients ne pouvaient être obtenues qu'à une granularité grossière (e.g., mensuellement ou bi-mensuellement), de nous jours, grâce aux infrastructures de comptage avancé (ou compteurs intelligents ou smart meters), les compagnies d'électricité peuvent les obtenir en temps réel. En tirant parti des données de smart meters, nous developpons un framework versatile pour segmenter les clients, suivons les changements des clusters à travers le temps et identifions les caractéristiques clés qui définissent un cluster.

Additionnellement, bien que la consommation au niveau résidentiel est difficilement prévisible, il peut être utilisé pour améliorer les prévisions au niveau agrégé en segmentant d'abord les ménages en plusiers clusters, puis en prédisant la consommation d'énergie de chaque cluster et enfin en agrégeant ces prévisions. Les améliorations fournies par cette stratégie dépendent non seulement du nombre de clusters, mais aussi de la taille de la clientèle desservie. En outre, nous developpons une approche pour modéliser l'incertitude de la demande future. En contraste avec les travaux antérieurs qui utilisaient des méthodes de calculs coteuses, comme la simulation, le bootstrapping ou l'ensemble, nous construisons des intervalles de prédiction utilisant directement la moyenne conditionnelle et la variance, variables dans le temps, de la demande future.

Bien que l'analyse des données de consommation d'énergie des clients soient essentielles pour comprendre les comportements des clients, elle peut aussi violer leur vie privée, avec tous les risques y attenant. La première partie de cette thèse se conclut par l'exploration de représentations symboliques des données de smart meter qui permettent aux algorithmes d'apprentissages de les utiliser similairement. Celles-ci présentent donc un compromis entre analyse précise et la protection de la vie privée des clients.

Dans la seconde partie de cette thèse, nous nous concentrons sur des mécanismes pour inciter des changements dans l'utilisation d'énergie par les clients dans le but de maintenair la stabilité du réseau électrique, i.e. dans le cadre de la réaction à la demande (Demand Response ou DR). Nous complétons le travail antérieur dans ce domaine (qui ciblait généralement de grands clients industriels) en étudiant l'application de DR aux clients résidentiels. Nous étudions tout d'abord l'influence de données de référence de DR, i.e. des estimations de la consommation des clients en l'absence d'événement de DR. Alors que la littérature à ce jour s'est concentrée sur la précision et le biais de référence, nous avous avons dépassé ces concepts en expliquant comment une valeur de référence affecte la participation du client à un événement de DR et comment ceci impacte tant le client, que le profit de la compagnie. Nous poursuivons par la présentation d'une stratégie pour accorder la demande à l'offre en utilisant une système d'enchères à multiples unités accomplies par des agents intelligents au nom des clients. Cette thèse se conclut par l'obtention de mesures incitatives comportementales issues de la clientèle pour promouvoir et maintenir la participation des clients dans les programmes de DR.

Mots-clés : smart grid, analyse de données, segmentation des clients, smart meter, prévisions de charge, vie privée, demand response, durabilité, tarification dynamique, système multi-agents, comportement humain, crowdsourcing

Acknowledgements

First and foremost, I would like to thank my advisor, Karl. This thesis would not exist without him. In the admission interview, I told him: “but, I do not have the background (e.g., P2P) that your lab has.” He replied, “It’s ok. That, you can learn. The more important things are your attitude and your potential.” Since then, I knew that I will try my best, day and night, to not let him down. I am very grateful that he always believes in me. He gives me enough freedom to take initiative on my research. I like that very much. He has vast experiences and is very good at what he is doing. He could look into the paper that I have written, spotted a flaw, told me about it, and (after fixing it) it eventually became one of the paper strengths. Sometimes it took me days before I could digest what Karl said. He is able to see far away things that did not even cross my mind. And I think that it’s cool! Lately, do not want to miss his wisdom, I asked his permission to record our meetings. He agreed. Then, of course, I needed to play it several times before I understand them $\ddot\smile$ In short, I cannot ask for a better advisor.

I would like to thank my thesis committee: Prof. Faltings, Prof. Le Boudec, Prof. Pedersen, and Mathieu. I can’t thank them enough for their precious time and availability. Especially to Prof. Pedersen and Mathieu who had to travel for attending my defense. They are leaders in their fields and have a lot of experiences as well. Every single discussion with them is sharp and constructive.

I would like to thank Matteo. It is hard to imagine completing this thesis without him. Working with him has been a privilege. He is a genuinely nice person and great colleague. Sharing similar academic background and interest, our collaboration simply ‘click.’ Everything moves very fast. We also challenge each other ideas at times, but in positive and constructive ways. Our collaboration is not an addition, it is a multiplication.

I would like to thank the Wattalyst consortium. To Thanasis, Dipanjan, and Zhixian that helped me to ‘take off’ in my first year. To Deva and Tanuja who hosted me in their Smarter Energy group in IBM Research India, summer 2013. The collaboration with them has been very fruitful and is the heart of Chapter 3. And to other partners: Mikael, Arne, and Jan from LTU, Achim from RWTH Aachen, George S., George T., and Marilena from AUEB, Iñiaki and Mikel from Tecnalia, and Pau, Carlos and Maria from Sampol, our face to face meetings have always been insightful and full of joy. Wattalyst has ‘raised’ me. From knowing almost nothing, I grew and learned from every discussion we made.

I would like to thank my colleagues and friends during my internship at IBM Research Ireland, early 2014. Especially to Mathieu (again $\ddot\smile$) and Bei for their unlimited patience. They often have to explain some basic concepts to me (because somehow I have never heard of them) until it became crystal clear to me. All good things in Chapter 5 is because of them. To Hussein, coming from the same institution, EPFL, we only met each other at IBM Research Ireland (both of us were interns at that time). I learned a lot from him, especially the kind of basic/fundamental (machine learning related) questions that you always wanted to know but were afraid to ask.

He explained them effortlessly. To Olivier who introduced me to the CER dataset, in summer 2012. The dataset soon became my favorite. In addition to Chapter 3, 4, and 7 in this thesis, the dataset has also contributed to several other publications that I co-authored.

I would like to thank Kate Larson. She was very instrumental in introducing me to multiagent system. It is hard to imagine completing Chapter 8 without her.

I would like to thank Julien. We have several collaborations, where one of them resulted in Chapter 6. A very talented friend. Being one of the two Swiss in our lab (surprisingly small number, btw), I abused him with lots of questions about Swiss.

I would like to thank Samuel, a smart student whose work paved the way for Chapter 4.

I would like to thank Ulysse Rosselet for introducing the open innovation crowdsourcing platform, Atizo. He also helped me to set up the crowdsourcing experiment in Chapter 9.

I would like to thank Hông-Ân for helping me with the French abstract. Together with Christoph, their knowledge about SBB and wine collection amazes me.

I would like to thank Zico Kolter. We have never met, but Chapter 1 is inspired by his CMU lecture *Computational Methods for the Smart Grid*.

I would like to thank Darren for proofreading part of this thesis.

I would like to thank Chantal for her support in administrative tasks. I asked her a lot of questions, from formal to informal, from regulations to social norm, about our lab, EPFL, Lausanne, and Switzerland. Working at the EPFL without her would be very hard.

I would like to thank all my friends. To Waheed who guided me to EPFL. Being friends since we were in Dresden, I practically follow his path to go to EPFL. To Myriam, Aleksandra, and Khoa, I cannot ask for better flatmates. I am very lucky to have them. To Fauziah, mbak Henny, mbak Rini, Ihan, Yolanda, and Aufar. Every moment I spent with them is simply a joy. Not only because I can freely speak Indonesian, but also because I can act and eat Indonesian. To my office mate, Alexandra. I learned a lot from her. About academics, humanity, and ... disaster (yep, this is what she cares about now $\ddot\smile$). To Mehdi and Pamela, I remember as if it was yesterday, we sit together at the top of the BC building, together with Alexandra, enjoying the warm sunny weather and eating together during the MICS summer school 2011, prior to joining EPFL. To Hoyoung, Zoltan, Sofiane, Rammohan, Surender, Saket, Hung, Michele, Tian, Jean-Eudes, Jean-Paul, Rameez, Hamza, Hao, Martin, Berker, Alevtina, Sarvenaz, Tomasso, Tam, Amit, Iuliia, Burak, Nataliya, and Imen. To Amitabh & Mohita, Jiaqing, Mihai, Florin, Lauren, and Jasmina. To all friends in EDIC. Without them, life in EPFL and Switzerland would only be half as good.

To Andrea, for her constant love and support that are much bigger than I can imagine. I am deeply thankful and grateful for her presence in my life. Stars have been much brighter since then. Also to her family for their encouragement. It's always fun to be near them.

And the last but not the least, to my family, who always believe in me. No words can express my gratitude towards them. Especially to my mom, my sister, and aunt Lily who never stop supporting me. I am very lucky to have them. They put their faith in me, letting me go more than ten thousand kilometers away to chase my dream.

Contents

List of Figures

List of Tables

Introduction

1

1.1 Developing Sustainable Energy Systems

1.1.1 Today's Energy Challenges

Sustainable development is defined by the United Nations as development that meets the needs of the present without compromising the ability of future generations to meet their own needs [World Commission on Environment and Development, 1987]. World energy consumption, however, has increased rapidly since the industrial revolution in the 1800s (see Figure 1.1) Its cause is not only the increasing world population, but also the increasing energy consumption per capita (Figure 1.2). Unfortunately, more than 80% of humanity's energy consumption is made possible

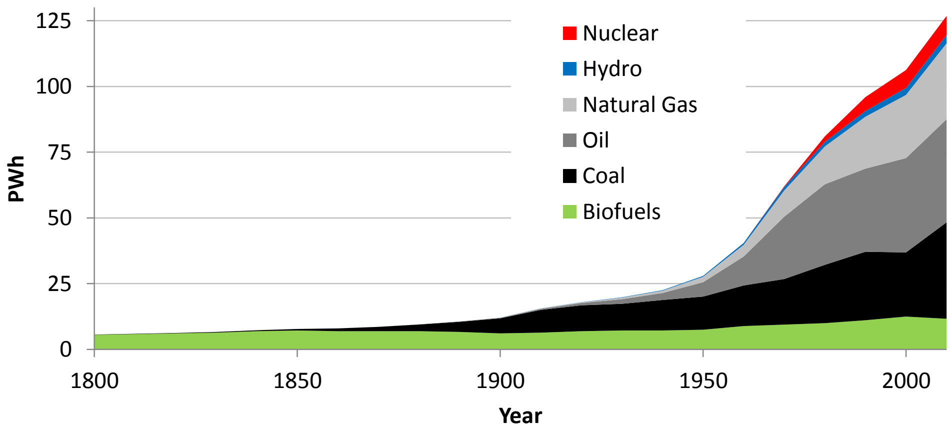
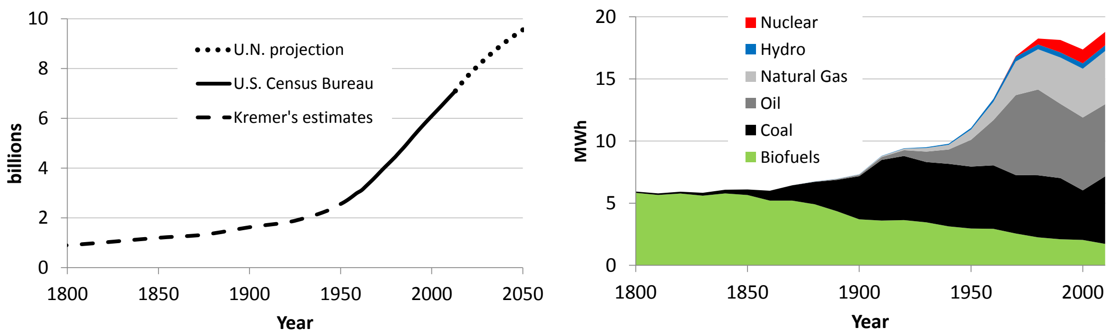
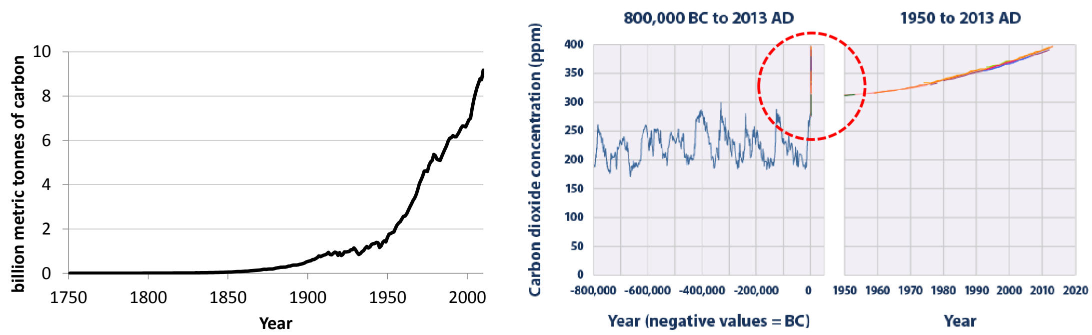


Figure 1.1: World annual energy consumption in petawatt-hours, i.e., 10^{15} Watt-hours [Roser, 2014b]. We converted the original figure, in Exajoules, to Watt-hours so as to be consistent with the rest of the thesis. This figure does not show the share of renewable energy sources such as wind and solar, which is approximately less than half that of hydro power's: hydro power constitutes around 2.4% of the world energy mix; other renewable energy sources currently amount to only 1.1% [International Energy Agency, 2014].



(a) (b)

Figure 1.2: (a) World population estimate [Kremer, 1993, Roser, 2014c, United Nations, 2014, U.S. Census Bureau, 2014]; (b) World annual energy consumption per capita—computed by dividing the total world energy consumption in Figure 1.1 by the world population estimate.



(a) (b)

Figure 1.3: (a) World CO_2 emission [Roser, 2014a]; (b) Global atmospheric concentrations of carbon dioxide over time [U.S. Environmental Protection Agency, 2014].

by burning fossil fuels, and this is at the heart of at least two major issues. The first is that fossil fuels are non-renewable energy resources: their use is unsustainable as mankind is consuming them at a much faster rate than nature might reproducing them. Despite this, there might still be enough fossil fuels for the next century. The second major issue, and the most pressing, is climate change.

Since the industrial revolution, CO_2 emission has increased dramatically (see Figure 1.3a). If we go further back in time, although CO_2 concentrations in the atmosphere fluctuate over periods of hundreds of thousands of years, the current concentration (approx. 400 ppm) is much higher than the average has ever been in at least 800,000 years (Figure 1.3b). Humanity might actually be degrading the condition of the planet such that it will be more costly to fix the situation later than it would be to fix it now. CO_2 is a green-house gas, i.e., its increasing concentration in the atmosphere increases the planet's surface air temperature. A doubling of the CO_2 concentration would increase that temperature by a constant amount, also referred to as the *equilibrium climate sensitivity* (ECS). The Intergovernmental Panel on Climate Change (IPCC) estimates that ECS is likely to range from 1.5°C to 4.5°C [Bindoff et al., 2013].

From energy in general, let us now move our focus to electricity. While it accounts for only about 18% of humanity's current energy needs, it is becoming increasingly crucial in everyday life [International Energy Agency, 2014]. Additionally, its share in the energy mix is likely to increase, in part due to the growing adoption of electric vehicles (which is also part of the effort to reduce our carbon emission). Although it is not always most efficient to use electricity,[1] it is a high-grade energy that can be easily converted to other forms of energy. Thus, if humanity can produce electricity sustainably, it can fulfill its energy demands sustainably as well. In other words, generating electricity (and fulfilling energy demands) using renewable energy sources such as wind or solar power[2], might be just the recipe for sustainable energy systems.

1.1.2 The Smart Grid

On the one hand, energy demand is ever growing. The increasing market penetration of electric vehicles (EVs), as an effort to reduce carbon emission, also poses a new challenge to electricity grids as EVs draw a large amount of energy in a very short time. Charging one EV battery can consume up to 32 kWh (comparable to one household's daily consumption) in just a few hours [Hess et al., 2012, Ramchurn et al., 2012]. On the other hand, most current energy demand is fulfilled by burning fossil fuels, which is not sustainable. To move towards sustainability, therefore, policy makers must encourage and incentivize power generation from renewable energy sources, such as wind and solar power. However, power generation from these energy sources is intermittent and unpredictable, which adds additional complexity to efforts at balancing supply and demand.[3]

Solving these challenges will require fundamental changes to today's grids, which are based on 40-year-old technology. This necessity has stimulated the creation of the so-called smart grid, i.e., *a fully automated power delivery network that monitors and controls every customer and node, ensuring a two-way flow of electricity and information between the power plant and the appliance, and all points in between. Its distributed intelligence, coupled with broadband communications and automated control systems, enables real-time market transactions and seamless interfaces among people, buildings, industrial plants, generation facilities, and the electric network* [U.S. Department of Energy, 2003]. Realizing this vision will require several key components, as follows.

- **Bidirectional energy flow.** In addition to the (conventional) energy flow from suppliers to customers, smart grids also enable distributed generation (DG) by allowing energy to flow in the reverse direction. That involves some customers generating energy locally (which then makes them prosumers), e.g., by using rooftop solar panels and injecting energy back into the grid when needed, thus reducing stress on the grid.

- **Bidirectional information flow.** While there is almost no information flow in traditional grids, in the smart grids, information flows from utility companies to customers in the form of pricing or other control signals, and from customers to utility companies in the form of energy consumption measurements (e.g., using advanced metering infrastructure or smart meters).

[1] Using gas directly for heating a home might be more efficient than using electricity. For instance, in gas-turbine power plants, gas is burned to heat a boiler and steam turns the turbine; the generated electricity is then transmitted through transmission/distribution lines with some losses, before it is finally used to power electric heating at home.

[2] Hydro power is also a of renewable energy resources, however, its potential for development is limited compared to that of solar or wind power.

[3] Electricity generation and consumption must be balanced across the entire grid at all times, otherwise the grid risks collapse, resulting in a blackout.

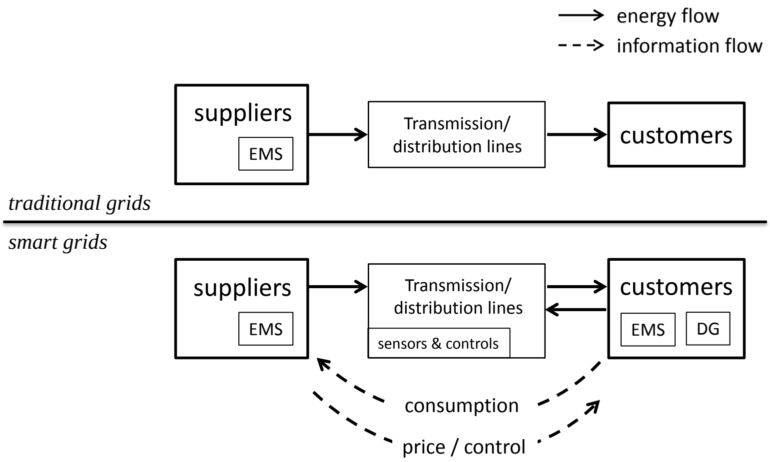


Figure 1.4: A simplified illustration of how key grid components relate to suppliers (and/or utility companies) and customers. EMS, energy management systems. DG, distributed generation.

- **Advanced sensing and control.** In a smart grid, power flow sensors and controls are placed throughout the transmission/distribution lines to ensure system stability, i.e., to quickly detect energy theft/sabotage, isolate failures before cascades into major blackouts, and guarantee uninterrupted services by rerouting energy transmissions while the problem is physically repaired by line technicians.

- **Energy management systems** on both the supply and demand sides. While decision support and control mechanisms are required on the supply side to effectively and efficiently operate the grid, customers (the demand side) want to minimize their energy bills (e.g., in the presence of dynamic pricing or other incentives) whilst maximizing their comfort.

Figure 1.4 illustrates how various key components relate to suppliers (and/or utility companies) and customers in smart grids in comparison to traditional grids. When they are put in place, smart grids are capable of [EU Commission Task Force for Smart Grids, 2010]:

- better facilitating the connection and operation of generators of all sizes and technologies,
- significantly reducing the whole electricity supply system's environmental impact,
- ensuring system reliability, quality, and security of supply,
- more efficiently maintaining existing services,
- providing customers with greater information and options on their energy use, and
- allowing customers to play a more active role in optimizing the system's operation.

Traditionally, when it came to balancing supply and demand in electricity systems, it was always the supply side that had to match the demand. However, if we wish to increase the share of renewable energy sources (which are both intermittent and unpredictable) in our energy supply mix and hence build sustainable energy systems, it might not always be possible for the supply side to follow demand (see, e.g., Figure 1.5). Thus, in the absence of appropriate energy

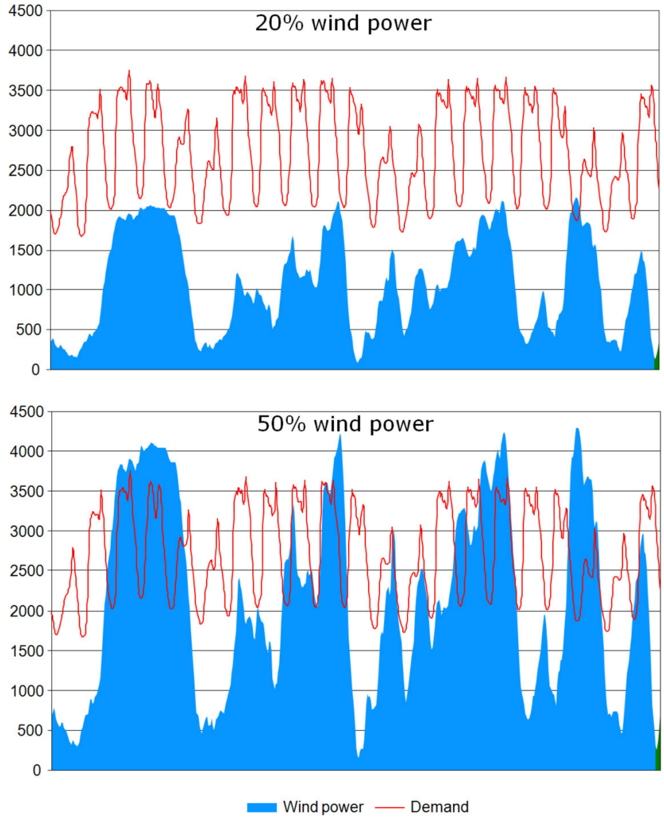


Figure 1.5: A sample (3 weeks) of wind power production and electricity demand in western Denmark [Kok, 2013]. The top figure is the situation in 2008, where 20% of total demand was covered by wind power. The bottom figure is the expected situation in 2025 where wind power is projected to cover 50% of demand. Note that, the production curve rarely matches the demand.

storage technologies,[4] the demand side will have to be managed in order to match the available supply. The emergence of smart grids and their capabilities, as outlined above (especially the last two points) are the starting point in any discussion of demand-side participation in the creation of sustainable energy systems.

1.1.3 Demand-Side Participation

Encouraging behavioral change or altering customers' energy consumption for the benefit of the whole energy system has long been known as *demand-side management* (DSM). It comprises *energy efficiency* (EE) and *demand response* (DR). EE includes all permanent changes, such as exchanging old, incandescent light bulbs for compact fluorescent lamps, upgrading inefficient

[4] It is also worth mentioning here that the field of energy storage is constantly seeking scalable and efficient technology to smooth out the intermittent and unpredictable nature of renewable energy sources, i.e., by storing the energy produced when supply exceeds demand and releasing it when supply falls below demand. One increasingly popular direction for research involves increasing the capacity of lithium-ion batteries (which dominate recent EV battery technologies) while reducing their cost [Scott, 2014]. Another involves reinventing compressed-air storage (see LightSail [Fong, 2014], Hydrostor [Kumagai, 2014], or an EPFL startup – Enairys [Lemofouet, 2014]).

ventilation systems, and improving buildings' thermal properties, e.g., by installing additional insulation [Palensky and Dietrich, 2011]. EE results in permanent (and constant) savings in emissions and energy use. DR includes *customers' changes to their normal electricity consumption patterns in response to changes in the electricity price, or to incentive payments designed to induce lower* (or higher) *electricity use at times of higher* (or lower) *wholesale market prices or at times when system reliability is jeopardized* [U.S. Department of Energy, 2006]. Although EE is always welcome, DR is attracting more and more attention from both researchers and policy makers, since it opens up possibilities for altering demand when it is most needed (and it may well induce greater demand reductions than EE), resulting in more aggressive savings at times. There are two types of DR programs.

(i) **Price-based DR** programs expose customers to dynamic pricing (instead of the more commonly used flat-pricing) depending on the time of day or other factors (e.g., market conditions). Customers are thus expected to lower their consumption when the price is high. Some examples of these programs includes the following.

  - *Time-of-use pricing (TOU)* divides the day into several time blocks (e.g., peak, off-peak) and assign different prices to each time block.
  - *Critical-peak pricing (CPP)* uses TOU as a basis and replaces the peak price with a much higher price when a specific condition is triggered, e.g., when grid stability is at risk, or when the market price is much higher than usual.
  - *Real-time pricing (RTP)* exposes customers to prices that fluctuate hourly, reflecting market conditions. Prices are announced on a day- or hour-ahead basis.

(ii) **Incentive-based DR** programs provide incentives (e.g., bill rebates, redeemable vouchers, or other benefits) to reduce (or increase) consumption at specific times (called *DR events*) requested by the program owner, either due to alarming grid conditions or market prices. Some of these programs are as follows.

  - *Direct load control* allows the program owner to shut down customers' electrical appliances such as washing machines, air conditioners, or water heaters, remotely.
  - *Interruptible service* offers customers special (cheaper) tariffs for agreeing to reduce their consumption during a DR event. The frequency of events is agreed beforehand. Failure to reduce consumption might incur penalties.
  - *Demand bidding/buyback programs* allow customers to provide offers (or bids) to reduce a specific amount of load during a DR event.
  - *Emergency DR programs* offer certain incentives to customers who reduce their loads during a DR event.
  - *Capacity market programs* allow customers to offer load reductions, as opposed to generating extra electricity, in capacity markets. Customers are paid in the form of reservation payments and other payments depending on the amount of load requested by the market. Failure to deliver the reduction incurs penalties.
  - *Ancillary service market programs* allow customers to bid load reduction for providing ancillary services in the market. If their bids are accepted, they are paid for being on standby and receive an additional payment when their energy use reduction are needed. Similarly to capacity market programs, failure to reduce consumption incurs penalties.

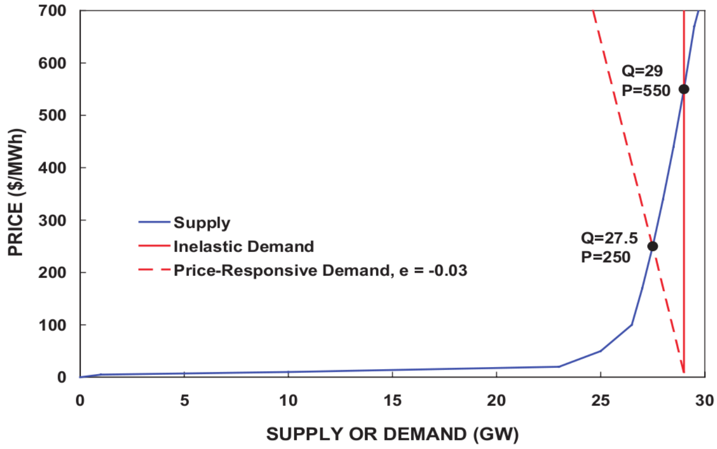


Figure 1.6: An illustration of supply and demand curves during the California energy crisis in summer 2000 [Hirst and Kirby, 2001]. P, price. Q, demand. The market clears at 29 GW for \$550/MWh. If consumers are even modestly sensitive to prices, the market could clear at 27.5 GW for \$250/MWh. The dashed line is a demand curve with a price elasticity of about 0.03.

The reliable operation of electricity grids necessitates a perfect balance between supply and demand in real time. While it has traditionally always been the supply side that followed the demand side, as we outlined in the previous section, the increasing market penetration of renewable energy sources means that it may no longer be possible for this to continue. If this were the case, DR promotes demand-side efforts to match the available supply. The demand side could make these efforts in several ways, e.g., reducing (or increasing) their consumption during a DR event, shifting their consumption during a DR event to other time slots (and vice versa), or drawing (or storing) energy from on-site generators (or storage devices) during a DR event.

In addition to supporting the future integration of renewable energy sources, DR is also currently useful for reducing electricity market price spikes that are typically caused by the high cost of running "peak" generators[5] during periods of very high demand (e.g., due to extreme weather conditions). Thus, demand reduction induced by DR could prevent the need to operate such expensive generators, lower the market price, and reduce carbon emission. For example, during the California energy crisis in summer 2000, the market cleared at 29 GW for \$550/MWh [Hirst and Kirby, 2001]. However, a 5% demand reduction would have reduced the price by 50% (see Figure 1.6). In Table 1.1, we summarize DR benefits and costs.

Some standards for automating DR, such as OpenADR,[6] have also been proposed. Simulations have also been built, e.g., PowerTAC [Ketter et al., 2014] which aims to find the best (pricing) strategy for energy retailers, and DRSim [Wijaya et al., 2013a] which aims to simulate customers' energy consumption. While DR could be one of the cheapest and greenest solutions to the current challenges faced by the electricity sector, it has been implemented only for large, industrial customers.[7] To maximize its full potential, DR implementation needs to be pervasive, including residential customers as well. Although, smart grids can provide the necessary infrastructure and technology, there are still some challenges to address in order to have

[5] These generators can be started and shut down quickly, but are more expensive to run and carbon intensive.
[6] http://www.openadr.org/
[7] With an exception to direct load control, which has also been offered to residential customers.

Table 1.1: A summary of DR program benefits and costs [Albadi and El-Saadany, 2008].

	Customers	Program owners
Benefit	• bill savings • outage reduction • incentive payments	• price/cost reduction • capacity increase
Cost	• comfort reduction (e.g., due to consumption shifting/reduction) • enabling technology (e.g., smart thermostats, energy management systems, on-site generators, storage devices maintenance)	• advanced metering infrastructure • billing system upgrade • program administration and marketing

a pervasive and successful DR program (these challenges are also the focus of this thesis), i.e., (1) understanding customers energy consumption behavior, identifying the root causes of their consumption, and estimating their future demand, which would be useful elements for deciding whether a DR event should be launched and for determining which customers to target,[8] and (2) providing just the right incentives to encourage customer participation and achieve program targets (see also our review of various DR mechanisms in Section 2.2).

1.1.4 The Role of Computer Science

Although developing sustainable energy systems requires interdisciplinary contributions from various fields, such as physics, chemistry, engineering, and economics, some of its key features present challenges (and opportunities) that have long been the research focus within the computer science community.

Communication network and cybersecurity In contrast to the traditional grids where information flow is minimal, in smart grids various information flows from and to any nodes in the grids. It could be, e.g., measurements from various sensors, control signals from grid operators, or pricing signals from utility companies to customers. Thus, a secure, robust and reliable communication network is required. It must be protected and made resilient against failures and attacks. Strong encryption and authentication techniques also plays a key role in securing data transmission at any points in the network. Additionally, grid components of various types, models, and manufacturers require a set of standards to ensure end-to-end communication and data exchange [see also Bouhafs et al., 2012, Ipakchi and Albuyeh, 2009]

Data storage and computing platform Sensors placed throughout the grids (including smart meters in customer premises) will soon generate an enormous amount of data (hence, the "big data" buzz around smart grids).[9] It requires cost effective and scalable data storage and computing platforms. Thus, constant innovation in both, hardware and software, is essential. To this end, cloud computing services (e.g., platform as a service, software as a service) become

[8] Although estimation of supply availability in the future is also important before launching a DR event, it is beyond the scope of this thesis.

[9] See, e.g., [Byrne, 2011, Danahy, 2009, John, 2013, Rose, 2014].

increasingly popular solution since it enables utility companies (practically new actors in information technology business) to perform a seamless and elastic resource usage and acquisition.

Analytics Having all the data by itself is meaningless. Thus, the next step is turning it into an actionable insight, and realizing the *smart*ness of the grids. Making sense of smart grid data requires data integration, statistics and machine learning algorithms to discover knowledge, patterns and anomalies, and interface and visualization design to present the insight the quickest and in the most comprehensible form. Depending on the data origin, there are at least three types of analytics.

- *Supply analytics* make sense of the electricity production data. It is especially useful to estimate future supply of renewable energy sources, such as wind/solar power, which is highly dependent on weather conditions. In practice, it supports generation planning, unit commitment, and deciding whether a DR event needs to be launched.

- *Network analytics* make sense of the data from the transmission/distribution lines, e.g., voltage, current, phase, or power flow data. Analytics on this data is useful for network planning and improving grid security and reliability, such as identifying failure or anomaly early, and detecting energy theft and congestion in the grids.

- *Demand analytics* make sense of customer energy usage data and aim to understand customer consumption patterns, It can be used, for example, to estimate future demand, which together with supply analytics, can be leveraged for planning generation and buying energy from the day-ahead market (which is typically cheaper than the intra-day market) if necessary. Other applications include customer segmentation for developing tariff structure, and selecting the right customers to target in a DR event. Since the data contains customer sensitive information, customer privacy protection is also an important issue.[10]

Multi-agent systems Daily operation of smart grids involves various actors, such as the energy suppliers, grid operators, the energy market, and the consumers. To this end, an application of multi-agent systems is to study the interaction between these actors, by modeling them as (selfish, altruistic, or anywhere in between) agents. For example, by assuming specific (but preferably realistic) customer models, mechanism design[11] can be used to develop a DR mechanism to achieve a desired objective, e.g., flattening peak demands. Coupled with learning algorithms, a software agent resides in customer's premise (or a smart home agent) can be used to learn customer's preferences and react to pricing/incentive signals sent by a utility company to minimize energy cost while maximizing customer's comfort. Multi-agent systems could also be used to study the interaction among many of such agents (see also the discussion about the herding effect in Section 2.2.1), and the interaction between those agents and other actors in the grid (e.g., energy brokers [Ketter et al., 2014]).

[10]Energy usage data can be used, for example, by criminals to identify the best times for a burglary, by the press to capitalize on public interest in famous individuals' activities, by the government to monitor tax-specific activities, by insurance companies to adjust premiums or claims, and by almost any companies to deliver targeted advertisement based on customers' life style and appliance usage patterns.

[11]Also known as algorithmic game theory, or reverse game theory.

1.2 Thesis Scope and Contributions

This thesis focuses on demand analytics and leverages the flow of information available in smart grids. Furthermore, it considers the economic subsystem instead of the physical subsystem,[12] and assumes that no appropriate energy storage solutions are available to the supply side.

This thesis contributes to efforts to understand customers' energy consumption behaviors as follows.

Chapter 3 We introduce a versatile customer segmentation framework, track cluster changes over time, determine individuals who changes her behavior (hence, her cluster), and identify the key characteristics that constitute a cluster,

Chapter 4 We forecast the electricity demand of residential customers. Although it is a hard problem due to the irregularities in household energy consumption, it can be leveraged to improve aggregate forecasting using the Cluster-based Aggregate Forecasting (CBAF) strategy.[13] We find that the improvement provided by CBAF depends not only on the number of clusters but also on the size of the customer base: the CBAF strategy only provides an improvement when the size of the customer base is greater than a certain threshold.

Chapter 5 In addition to point forecasts (as in Chapter 4), we also forecast the uncertainty in electricity demand. While in the literature to date prediction intervals were typically developed using computationally expensive approaches such as bootstrapping or ensemble, we introduce a method to construct them directly using time-varying conditional mean and variance. We also introduce an online learning algorithm to adapt the prediction intervals to the non-stationary nature of electricity demand.

Chapter 6 We reduce the data size and privacy risk of smart meter data by converting it into symbols while still allowing various analytics (such as classification or forecasting) to be performed on top of it.

This thesis also contributes to efforts to develop DR mechanisms to incentivize and alter customers' energy consumption behaviors as follows.

Chapter 7 We evaluate not only the accuracy and bias of DR baselines but also their impact on stakeholders' profits. We also show that more positively biased baselines foster greater customer participation.

Chapter 8 To reduce high costs due to peak-hour demands, rather than relying on customers' willingness to act, we explicitly cut peak demand and match the demand side to the available supply using multiunit auction. Furthermore, the auction can also be used to match the demand side with a supply curve of any shape (e.g., due to the high market penetration of renewable energy sources), as long as it fulfils the minimal load guarantee.

Chapter 9 We perform a crowdsourcing experiment to elicit effective behavioral incentives for residential customers. Using Fogg's Behavioral Model, we classify and analyze the

[12] While the *physical* subsystem considers hardware that physically produces and transmits electricity, the *economic* subsystem considers the actors that are involved in the production, trade, or consumption of electricity and their mutual relationships [adapted from De Vries, 2004, Kok, 2013].

[13] Previous works also refer to this strategy as *dissagregate forecasting*, despite it actually forecasts aggregate demand. To avoid further confusion with individual and aggregate forecasting, we use the term Cluster-Based Aggregate Forecasting.

submitted ideas. Additionally, we show how several submissions can be combined into a more complete solution aimed at sustaining customer participation.

2 State of the Art

This chapter reviews the state of the art of demand analytics (i.e., customer segmentation and load forecasting) and various demand response (DR) mechanisms. While demand analytics could help to determine when to launch a DR event and which customers to target, research in DR mechanism enable utility companies to implement the most appropriate incentives to alter energy consumption behavior of their customers.

2.1 Demand Analytics

This section surveys the state of the art of two demand analytics tasks: customer segmentation and load forecasting.

2.1.1 Customer Segmentation

Psychographic Segmentation

Psychographic segmentation aim to segment people in terms of how they think, feel and act. The study is typically done by completing a survey customers containing attitudinal and behavioral questions. Pedersen [2008] perform a psychographic segmentation to find customer groups based on their behaviors and attitudes toward electricity and conversation. The behaviors and attitudes are further detailed into several categories, such as how they use lighting, plug-in device, dishwashing, laundry, space heating/cooling, and water. He find that the customers can be divided into six customer groups, i.e., (from better to worse conservation ethic) tuned-out & carefree, stumbling proponents, comfort seekers, entrenched libertarian, cost-conscious practitioners, and devoted conservationists. The result is then used by a utility company (whose customers are studied by Pedersen) for its eNewsletter campaign and conservation programs, such as selecting a group of customers to be conservation role models in their community.

In addition to customer behaviors and attitudes toward energy conservation, Sütterlin et al. [2011] also survey customer attitudes in purchasing new cars, their attitudes toward using public transport or their own cars, and their acceptance to public policy, such as renewing old power plants, or increasing price for appliances/cars with high energy consumption. As a consequence, while Pedersen's segmentation is more about how people conserve energy, Sütterlin et al. also take into account customer attitudes in purchasing energy-saving appliances or goods. They characterize customers into six segments: idealistic energy-savers, selfless inconsequent energy-savers, thrifty energy-savers, materialistic energy consumers, convenience-oriented indifferent energy consumers, problem-aware well-being-oriented energy consumers. The existence of the selfless inconsequent energy-savers is particularly interesting since they are inconsequent in translating their thinking into action, i.e., they are highly aware of energy conversation problem but less energy aware in their purchasing decision.[1]

In contrast to previous works, Sanquist et al. focus on how people consume energy, e.g., times per week oven is used, dishwasher loads per week, hours per week TV/computer is on, total hours light are on per day, and the size of air-conditioned (AC) area in the house [Sanquist et al., 2012]. Additionally, they also develop segmentation based on where customers live, i.e., city, town, suburb, and rural area. They find that people who live in the city consume the least amount of energy due to the low energy consumption for AC and laundry. Despite the low use of AC, people who live in the rural area consume the largest amount of energy due to the high energy consumption for laundry .

Load Pattern Segmentation

Psychographic segmentation relies on customers' answers on a survey. However, customers' answers might not actually in line with what they do [Peattie, 2001]. To this end, load pattern segmentation aims to cluster customers based on how they consume energy in reality, i.e., based on the metered consumption.

Segmenting Commercial and Industrial Customers There is a large body of literature on customer segmentation using load patterns of commercial and industrial customers to improve tariff structures (see e.g., [Chen et al., 1997, Chicco et al., 2003, Figueiredo et al., 2005, Kitayama et al., 2002, Ramos and Vale, 2008, Ramos et al., 2007, Tsekouras et al., 2007]). Chen et al. [1997] perform the segmentation simply based on customer contractual data, i.e., customer activity type (commercial or industrial) and voltage level (low, high, or extra high). However, Chicco et al. [2003] shows that grouping customers based on their contractual data might be ineffective in characterizing their electricity consumption behavior.

Therefore, several studies propose features that are derived from customer consumption data. Chicco et al. [2003], Figueiredo et al. [2005], and Ramos and Vale [2008] propose features based on statistics on customer daily consumption, such as the average, minimum, or maximum power demand during the day,[2] night impact, and lunch impact. Night impact is defined as the ratio between the average consumption during the night (23:00–06:00) and the day (06:00–23:00), while lunch impact is defined as the the ratio between the lunch time (12:00–14:00) and the day (06:00–23:00). In contrast, Ramos et al. [2007], Tsekouras et al. [2007], and Chicco [2012] propose features that represents customer typical daily load curve. More specifically, they divide the day into T time slots and define a feature vector of length T, where the i^{th} element is the customer typical consumption at the i^{th} time slot.

Several unsupervised learning algorithms have also been employed, such as kMeans [Tsekouras et al., 2007], hierarchical clustering [Ramos et al., 2007], and Self Organizing Maps [Figueiredo et al., 2005]. After customer segments have been built and a new tariff structure has been pro-

[1] This fact also supports Peattie's observation that green purchasers are not necessarily the same person as green consumers, and vice versa [Peattie, 2001].

[2] The ratio among them can also be considered, such as average/maximum, minimum/maximum, or minimum/average.

posed, a supervised learning algorithm, e.g., decision tree, can also be trained on top of the clusters to classify new customers, and determine their tariffs [Figueiredo et al., 2005, Ramos and Vale, 2008, Ramos et al., 2007].

Segmenting Residential Customers Due to the recent deployment of smart meters, there have been a growing interest on customer segmentation that focuses on residential customers. Similar to the works described previously, Räsänen and Kolehmainen [2009] and Flath et al.[2012] perform a customer segmentation to improve customer tariff structure.

When a new customer join, a utility company typically associates her to a customer class based on her house type (e.g., detached, terraced), heating type (e.g., eletric heating or not), and activity type (e.g., spare-time cottage, agriculture residence). In other words, the utility company performs psychographic segmentation. To this end, Räsänen and Kolehmainen propose to segment customers based on the statistical features derived from their energy consumption, such as mean, standard deviation, skewness, kurtosis, etc. Räsänen and Kolehmainen then apply kMeans algorithm, and by using Index-of-Agreement,[3] they show that newly developed segments perform better than the segmentation originally developed by the utility company. Similar to Räsänen and Kolehmainen, Flath et al. also apply kMeans clustering algorithm. However, they use customer load curve as a features and aims to provide insight specifically into designing time of use pricing. They suggest four key steps: determining the number of time zones, identifying the starting time for each time zones, determining the price for each zones (one price could apply to several zones), and maximizing supplier profit (by deriving demand elasticity from field tests).

Segmenting Daily Load Curve In contrast to the works described above which aim to cluster customers, Pitt and Kirschen [1999], Cao et al. [2013], and Kwac et al. [2014] focus on clustering customers' daily load curves. More specifically, given a set of (normalized) daily load curve collected from all customers they aim to group similar curves in the same cluster. Thus, it is possible that load curves of a customer belong to several clusters. Pitt and Kirshen apply decision tree clustering (using day, month, and load factor as explanatory variables), while Cao et al. employ hierarchical clustering, SOM, and kMeans, and Kwac et al. use adaptive kMeans (setting distance threshold parameter instead of k). Both, Cao et al. and Kwac et al. use customer daily load curve as features, and in contrast to the works presented up to this point (including Pitt and Kirshen's), they aim to support DR implementations and energy efficiency programs. For example, after clusters have been identified, utility companies could then classify the daily load curves of each customer, and (depending on the programs) target customers that have stable daily patterns (more predictable), or customers with peak demand at a particular time of day.

2.1.2 Load Forecasting

Long-term Forecasting

Long-term electricity load forecasting predicts electricity demand from one to several years ahead and is especially useful for planning capacity and networks (transmission/distribution). Mohamed and Bodger [2005] forecast annual national demand of New Zealand 16 years ahead. As features, they use gross domestic product (GDP), population, and electricity price. Linear regression is used to model the relationship between the features and the demand. To be able

[3] If P_t and O_t are the predicted and observed value at time t respectively, and n is the number of observations, then Index-of-Agreement is defined as $1 - \left(\sum_{t=1}^{n}(P_t - O_t)^2\right) / \left(\sum_{i=1}^{n}(|P'_t| + |O'_t|)^2\right)$, where $P'_t = P_i - \bar{O}$, $O'_t = O_t - \bar{O}$, and $\bar{O} = \frac{1}{n}\sum_{t=1}^{n} O_t$.

to forecast future values, each feature is first forecasted independently also using linear regression. Slightly different from Mohamed and Bodger, Amarawickrama and Hunt [2008] use GDP per capita, electricity price, and time trend to forecast Sri Lanka's electricity demand, 22 years ahead. Bianco et al. [2009] forecast Italy's demand 23 years ahead and use GDP, population, GDP per capita, and electricity price as features. They evaluate multiple regression models using various combination of features. Surprisingly, they find that the influence of electricity price to Italy's demand is negligible, and thus discouraging the use of pricing policies to promote energy-efficient programs in Italy.

Mid-term Forecasting

While long-term electricity load forecasting predicts demand one to several years ahead, mid-term forecasting predicts one to several months ahead. Pao et al. [2006] focus on Taiwan's monthly electricity consumption. Similar to the works in long-term forecasting, Pao et al. [2006] investigate the relationship of four economic factors: GDP, population, national income, and consumer price index. They use linear (regression) and nonlinear models (artificial neural networks) and find that national population and national income affect Taiwan's electricity consumption the most. Saab et al. [2001] forecast one-step ahead monthly electricity consumption of Lebanon. In contrast to Pao et al., they focus on univariate time series modeling and use Auto-Regressive (AR), and Auto-Regressive Integrated Moving Average (ARIMA).

While Pao et al. and Saab et al. both focus on monthly electricity load data, Chen et al. [2004] forecast daily maximum load of Eastern Slovakia for the next 31 days as part of the EUNITE competition.[4] Features available are historical half-hourly electricity demand from 1997 to 1998, average daily temperature from 1995 to 1998, and dates of holiday from 1997 to 1998. The task is then to predict the daily maximum load of January 1999. They apply Support Vector Regression, use day type (the 7 days of the week plus holiday) and daily maximum loads of the past seven days as features and train only on winter data (from October to March). Although temperature has been known to affect daily electricity demand (especially in the short-term forecasting, see the next section), they find that incorporating temperature as a feature does not increase the forecasting accuracy, due to the difficulty of predicting the temperature itself for the next 31 days.

Short-term Forecasting

Short-term electricity load forecasting predicts electricity demand from one hour to several days ahead and is typically used for generation scheduling. Significant forecasting errors could lead to overly-conservative or overly-risky scheduling, and (possibly) incur economic penalties. More specifically, forecasting demand way higher than the actual results in starting too many units and unnecessarily high levels of reserves. Conversely, forecasting demand way lower than the actual results in failure to provide the necessary spinning reserves.

Forecasting Demand of a Country or Utility Service Area Papalexopoulos and Hesterberg [1990] use linear regression to forecast the next-day hourly electricity demand. They use the historical load data, holiday, day of the week, temperature, and time of year (which could also account for distinguishing seasons of the year) as features. In addition to temperature, Hyde and Hodnett [1997] include several other weather variables, such as sunshine duration, wind velocity, and humidity. While Papalexopoulos and Hesterberg build only one model, Hyde

[4] The EUNITE project url can be found at `http://www.eunite.org`, while the competition can be found at `http://neuron.tuke.sk/competition/`. Chen et al. eventually won the competition.

and Hodnett develop 12 regression models, one for each month. Hong [2010] employs multiple linear and polynomial regression to forecast the next-week hourly load. Due to this rather long forecasting horizon, Ružić et al. [2003] propose to first model the total daily demand before modeling the intra-day hourly load.

Hagan and Behr [1987] apply Seasonal ARIMA, where autocorrelation and partial autocorrelation functions are used to determine the order of the auto-regressive and moving average processes, respectively. Then, one day and one week seasonal part are added, and different models for each of the four seasons of the year are built. However, incorporating the non-linear effect of temperature in ARIMA is not straightforward. To this end, they use a polynomial regression to model the load-temperature relationship. In contrast, Cancelo et al. [2008] model the relationship using a piece-wise function by first breaking the load-temperature curve into several segments. To account for various electricity consumption behavior throughout the year, Hagan and Behr develop different models for each season in the year. As another alternative, Taylor [2010] suggests to incorporate the intra-year seasonality explicitly in the model. That is, he propose to use triple-seasonal ARIMA to account for daily, weekly and yearly seasonality in the demand.

Artificial Neural Network (ANN) is also a popular method for short-term electricity load forecasting [Hippert et al., 2001]. For instance, Khotanzad et al. [1997] use ANN based on multilayer perceptron and backpropagation learning. In addition to hourly historical load data, they incorporate day of the week, temperature and humidity. Then, demand for each hour of the day is modeled separately using one ANN, resulting in total of 24 ANNs. Additionally, Sapankevych and Sankar [2009] outline several works that uses SVM for short-term load forecasting. Similar to previous approaches, typically temperature and humidity are also used as explanatory variables. Fan and Hyndman [2012] use Generalized Additive Models (GAM) to forecast half-hourly demand up to 7 days ahead, and develop a model for each 48 half-hourly period of the day. Furthermore, Ba et al. [2012] propose an online learning mechanism to adapt GAM's smoothing functions to the non-stationary nature of electricity demand.

Martínez-Álvarez et al. [2011] propose a forecasting method that is based on similarity of pattern sequences as follows.

(i) We first segment and cluster the data. The segment length depends on the forecasting horizon, e.g., daily. Thus, a segment can also be viewed as a daily curve. We then cluster the segments and assign a cluster label to each segment.

(ii) A pattern sequence (or label sequence) prior to the target day is extracted. The length of the sequence is a parameter that need to be defined.

(iii) We then search the sequence in the historical data.

(iv) The prediction result is the average of all segments immediately after the matched sequence.

In contrast to the previous approaches which use the real values of the time series, this method makes use only the segment labels for prediction. While Martínez-Álvarez et al. use kMeans algorithm for the clustering step, Shen et al. [2013] improve the algorithm accuracy by considering an ensemble of clustering algorithm to improve the forecast accuracy.

Recently, there is also competition dedicated to short-term forecasting for large demand, i.e., the Global Energy Forecasting Competition 2012 [Hong et al., 2014a, Nedellec et al., 2014, Taieb and Hyndman, 2014].[5] The competition focuses on hierarchical forecasting, where participants are required to forecast the hourly load of 20 zones and hourly system load of a utility company

[5]The competition also has another track for wind power forecasting.

up to 7 days ahead. The sum of the load over all zones at any given time should be equal to the system load. The competition also provides the participant with a list of public holiday and temperature data from 11 weather station.

Forecasting Demand of Residential Customers Due to the recent deployment of smart meters, forecasting energy demand at the residential level is a relatively new area. Ghofrani et al. [2011] consider only one household and very short forecast horizon, i.e., 15, 30, and 60 minutes using one day of training and one day of testing data. Veit et al. [2014] consider electricity consumption of two households with 14 and 44 days of measurements. They forecast from 15 minutes to 24 hours ahead and consider sampling frequency of 15 to 60 minutes, and find that ARIMA and ANN do not necessarily outperform the simple persistence forecast. However, one need to be careful in interpreting this result since Veit et al. perform only one-step ahead forecast, and then forecast longer horizons recursively. That is, for a forecasting horizon n, they use the historical data up to time t as the training data to forecast the value at time $t+1$, add it to the training data to forecast the value of time $t+2$, and continue in similar manner until they obtain the forecast for time $t+n$. The drawback of this approach is that the error from one time slot is propagated to the subsequent time slots. An alternative approach would be (which is also more common in the literature), to directly forecast the value at time $t+n$, i.e., by training a learning algorithm to forecast the value at time t' using only the historical data up to time $t'-n$, for all $t' \in [n, t]$.

Tidemann et al. [2013] evaluate forecasting accuracy of various learning algorithms at the transmission level (around 10,000 customers), distribution substation (around 150 customers), and individual customers. They find that forecasting error increases as the aggregation size decreases. Additionally, methods that works well to forecast demand at the transmission level does not necessarily work well for smaller aggregation size.

Chaouch [2014] applies a modified functional wavelet kernel (FWK) approach to forecast electricity demand at the household level. Let S_i denote the customer's load curve of day i. To forecast S_{d+1}, the method works as follows.

(i) Cluster the load curves up to day $d-1$, such that similar load curves belong to the same cluster.

(ii) Identify a cluster that is the most similar to S_d.

(iii) The forecasting result is the weighted average over all load curves in the cluster found in the step (ii) above. The weight of a curve is then defined by a kernel function on the discrete wavelet coefficient similarity between the curve and S_d.

The original FWK computes the result directly by weighted averaging over all curve,[6] and does not have the initial clustering step. Thus their approach, which is also called Cluster-based FWK (CFWK), can also be seen as the generalization of FWK, i.e., when we have only one cluster for all curves, then CFWK is essentially FWK. CFWK, however, took only the historical load curve as input, and therefore one need to modify it carefully to account for external factors, such as calendar variables or temperatures.[7]

2.2 Demand Response

This section reviews the state of the art of demand response (DR) mechanisms.

[6] The weight of a curve also depends on the similarity between the curve and S_d.

[7] Several works have also used demographic information to estimate electricity demand. See, e.g., [Jarrah Nezhad et al., 2014, Kolter and Ferreira, 2011, Mohamed and Bodger, 2005, Wijaya et al., 2014a].

2.2.1 Energy Consumption Scheduling

Several works have investigated DR through flexible scheduling of electrical appliances. Mohsenian-Rad et al. [2010], Li et al. [2011], Ramchurn et al. [2011], and Wijaya et al. [2013d] propose decentralized mechanisms using a software agent to find an optimal energy consumption schedule automatically. The mechanisms preserve customer's privacy since they do not need to reveal their detailed appliance schedule.

In their paper, Li et al. [2011] provide mathematical models to characterize the utility of various electrical appliances. They categorize residential appliances into several types as follows.

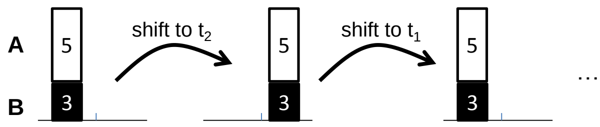
- Type 1: appliances that control the temperature in the building, such as heating or air conditioner.
- Type 2: appliances that bring utility when they completed a task before a certain time, such as electric vehicle (EV) battery charger, clothes washing machine, or dishwasher.
- Type 3: appliances that must be turned on for a certain period, such as light.
- Type 4: appliances that are used for entertainment, such as TV, games consoles, or computer. Utility from these appliances are assumed to be proportional to the total energy consumed.

To adjust customers' energy consumption behavior, they use dynamic pricing. The price is assumed to be a function of total demand, increasing and strictly convex. Instead of reducing peak loads (and thus flatten the demand curve), however, this pricing mechanism can lead to peak shifting due to the herding effect of customer's load-shifting behavior (see Figure 2.1 for illustration).

To address this problem, Mohsenian-Rad et al. [2010] propose a turn-taking mechanisms, which unfortunately does not scale well with the number of customers. Low et al. [2011] propose an iterative adjustment, i.e., in each iteration, customers update their energy consumption schedule only slightly toward the optimal schedule, depending on a stepsize constant. Ramchurn et al. [2011] address the problem by setting a uniform participation rates to the entire population. When a customer has a participation rate $p \in [0, 1]$, it means that she updated her schedule with a probability p. They find experimentally that p should be set reasonably small in order to avoid the herding effect. Wijaya et al. [2013d] validate this insight and continue further by investigating non-uniform participation rates and finding near-optimal (mixed) participation rates setting.

Instead of controlling price signal, Le Boudec and Tomozei propose a mechanism that controls the load signal, i.e., the rate at which a consumer may draw power from the grid [Le Boudec and Tomozei, 2011]. The mechanism avoids exposing price volatility (through dynamic pricing) to consumers by using flat pricing and introducing a service curve contract instead (agreed at subscription time). The contract consists of (1) the maximum power level a customer can draw at any time, and (2) a service curve that guarantees the minimum total energy a customer can draw at any given time window (see Figure 2.2 for example). A software agent is assumed to receive load control signals and schedule home appliances accordingly (to not consume energy more than what specified by the control signal).

Šikšnys et al. [2012] and Valsomatzis et al. [2014] assume that customers are able, and more importantly, willing to provide a so called *flex-offer*, i.e., a vector of energy consumption flexibility (the minimum and maximum level) over a time dimension. Flex-offers can also be generalized to unify the concept of supply and demand by introducing the notion of positive and negative flex-offers. Positive flex-offers refer to drawing energy from the grid (hence, the demand), and



	t_1	t_2	t_1	t_2	t_1	t_2
Load (L)	8	0	0	8	8	0
Total cost (L^2)	64	0	0	64	64	0
Cost per unit	8	0	0	8	8	0

Figure 2.1: An illustration of the herding effect [Wijaya et al., 2013d]. The cost depends on the squared load. The peak hour is initially at time slot t_1 (which costs 8 per unit). Afterwards, A and B shift their consumption to time t_2 (since it previously costs only 0 per unit). Rather than flattening the peak (and reducing cost), it causes a peak-shifting and makes t_2 the new peak. The peak load and total cost can be reduced if only A or B (but not both) makes the shift.

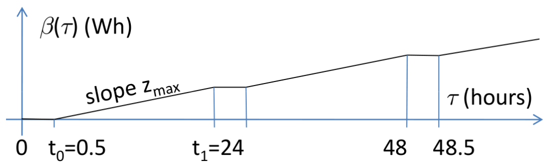


Figure 2.2: An example of a service curve [Le Boudec and Tomozei, 2011]. Intuitively, it allows a utility company to serve no power to a customer for at most 30 minutes in a day (or serve $(1-\frac{1}{x})z_{\max}$ watt for $x \cdot 30$ minutes, where $x \in [1, 48]$), and guarantee $z_{\max}$ watt for the rest of the day.

negative refer to injecting energy back to the grid (hence, the supply). There is also the third category, i.e., flex-offers that contain both, positive and negative values (mixed flex-offers), which refer to prosumers. Thus, the task of balancing the supply and the demand can be reduced into determining the value of each element in a flex-offer such as the sum of the values of all flex-offers for each time slot are as close as possible to 0. However, expecting customers to specify their flex-offers can be seen as a strong assumption.[8] To this end, Neupane et al. [2014] outline an initial look into automatically identifiying flex-offers for residential customers.

While the works we have described are more theoretical in nature, Ganu et al. develop nPlug (see Figure 2.3), a device prototype to control customer load according to grid conditions [Ganu et al., 2013]. The strength of nPlug is that it does not require any additional communication infrastructure or any changes to appliances or grids. It infers grid conditions by measuring voltage at a household level and schedules the load of the appliance attached to it. In its scheduling, nPlug takes into account customer constraints[9] and current grid conditions (i.e., avoids scheduling appliances at peak time slots). Although nPlug is a promising prototype, in order to encourage widespread deployment it needs to be coupled with a robust economic incentive, a mechanism to provide benefits to those who install it (compared to those who do not).

2.2.2 Storage Devices

In the previous sections, customers become grid friendly by adjusting her normal consumption schedule, thus providing some inconvenience. In this section, we discuss another DR mechanism

[8] For the supply side, however, providing flex-offers is to some extent similar providing unit commitments.
[9] Customers need to specify their scheduling constraints using nPlug's user interface.

Figure 2.3: nPlug prototype [Ganu et al., 2013].

that leverages storage devices (or battery) in customer premises. It aims to make the demand side grid friendly while minimizing their inconvenience (due to schedule changes) and energy bill at the price of the battery. Thus, this line of research typically focuses on finding battery charging/discharging schedules in order to minimize customers' (or the whole systems') energy cost. The general idea is to charge the battery and consume energy from the grid during off-peak hours (cheaper price periods), and to discharge the energy from the battery and draw no (or only a minimal amount of) energy from the grid during peak hours (more expensive price periods). For example, Mishra et al. [2012] compute battery charging/discharging schedules by considering day-ahead market prices and forecasting the next-day consumption patterns.

However, when all customers charge their battery at the same time, the system will be exposed to the herding effect (similar to what we have described in Section 2.2.1 and Figure 2.1). That is, rather than flattening the demand, the peak is shifted to the previously off-peak hours (see also [Carpenter et al., 2012]). To address this problem, Vytelingum et al. [2010, 2011] assume the existence of software agents at customer premises to automate and optimize the charging cycle, minimizing overall energy cost. The herding effect can then be avoided if agents' learning rate is set to a small constant (between 0.05 and 0.20).[10] They also provide a game-theoretic analysis, define a scheduling game, compute the Nash equilibria, and show that there is a specific participation rate (the proportion of customers that adopt storage) for which the equilibria can be achieved. Additionally, Le Boudec and Tomozei [2012] propose the use of an energy storage device to maintain customer comfort under a service curve contract,[11] find a charging schedule that satisfy customer non-elastic load and determine the sufficient battery capacity to perform the schedule in an online manner.

2.2.3 Cooperatives

Previously, we have discussed about DR at the individual level. In this section, we describe another line of work that focuses on coalition among customers to form a stronger entity for DR (i.e., collaborative DR). While the result of DR at the individual level is hardly predictable,[12]

[10] While the number here is similar to [Ramchurn et al., 2011, Wijaya et al., 2013d], this learning rate setting is more similar to Li et al. [2011].

[11] We have also explained service curve contract briefly in Section 2.2.1.

[12] Except for the service curve—since it can also be thought of as a variant of direct load control. There is a difference, however. While direct load control in the literature refers to the case where utility companies directly control electrical appliances, in service curve contract utility companies send signals to their customers specifying

the performance of collaborative DR is more reliable and predictable, and thus more attractive for planning. Collaborative DR is typically performed by rewarding those who can maintain their pre-agreed (reduced) load level, and incurring penalties to those who cannot.

Kota et al. [2012] propose a formation of a cooperative from a set of customers. The cooperative has a continuous presence in the market to sell *negawatt* energy, i.e., rather than providing energy generation, it bids to provide energy reduction. The revenue received from the market is then distributed among the members proportional to their reduction. However a member can behave maliciously by over-consuming electricity for some periods: inflating her consumption baseline and receiving higher reward (because she is perceived to provide higher reduction). To discourage this malicious behavior, Kota et al. propose a randomized mechanisms that, for each trading slot, assigns each member to the *reducers set* with probability ≤ 0.5.[13] Only the members in the reducers set are expected to perform the reduction and eligible for the reward. Although the members that are not in the reducers set are not expected to reduce, they are also expected to not increase their consumption.

Akasiadis and Chalkiadakis [2013] propose a cooperative specifically for day-ahead demand response, to shift consumption from peak to off-peak hours. The cooperative is formed when the grid issues a special price for some off-peak hours that is lower than the normal price for those hours. However, the price is only granted only if the amount of energy shifted from the peak to those off-peak hours is greater than some threshold (which most likely could be shifted only by a group of customers, rather than a single customer). Thus, the existence of the lower price can also seen as an incentive for the cooperative to help the grid shifts an amount of load from peak hours so that it does not need to activate the more expensive "peak" generators (which would have otherwise been activated).

While works described previously focus on providing energy reduction to the market, Vasirani and Ossowski [2013] propose a new business model which consists of a smart load balancing group (a set of customers) and an aggregator that buys an electricity from the day-ahead market (see also Figure 2.4 for an illustration). The mechanism works as follows:

1. Customers in the cooperative (or group) estimate their next-day demand, and specify their scheduling constraints.

2. The aggregator performs a global optimization on the group schedule in order to minimize the overall energy cost.

3. The aggregator sends the optimized schedule back to the customers.

4. The aggregator buys the required energy from the day-ahead market.

The energy price for the customers in this group is slightly above the market price (to leave a small profit margin for the aggregator) but below the retail price. Customers are expected to commit to their schedule, and the aggregator incurs penalties to those who deviate. Experiments suggest that it is possible for the aggregator to obtain some profit and for the customers to pay less than what they would have paid if they do not belong to this group.

---

the maximum load customers can draw at that moment.

[13] The mechanism works only if the reward is less or equal than the retail price. While Kota et al. consider only the case of over-consuming members, it can also be extended to the case where members inflate their baseline by *shifting* their consumption from other time slots, i.e., *if a member is assigned to the reducers set with probability $\leq$ 0.5 and the reward is less or equal than the member's shifting cost then it is not profitable for the member to inflate her baseline.*

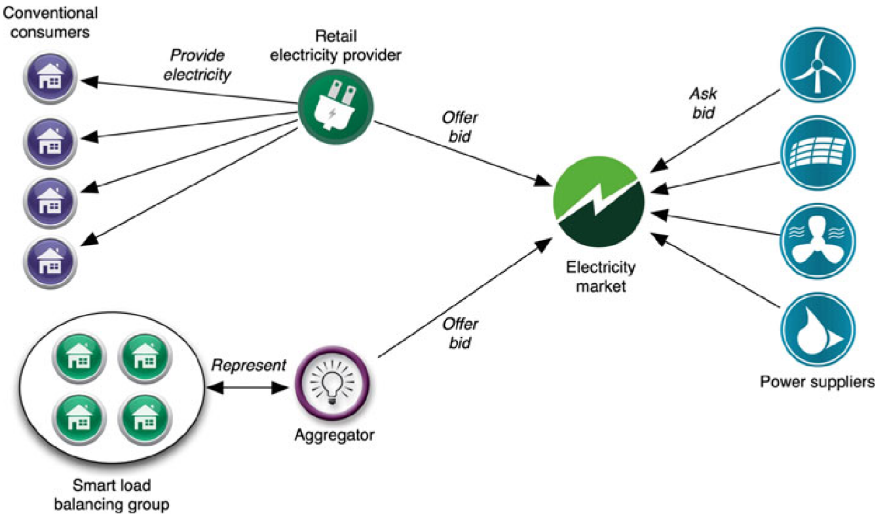


Figure 2.4: An illustration of the energy market structure with the existence of smart load balancing group [Vasirani and Ossowski, 2013]. From the market point of view, the aggregator is a single buyer (similar to other retailers).

2.2.4 Electric Vehicles

The large battery capacity of future fleets of electric vehicles (EVs) has been considered as a significant extra load that will have to be satisfied. Charging one EV battery can consume 32kWh (comparable to one household's daily consumption) in just a few hours [Hess et al., 2012, Ramchurn et al., 2012]. Additionally, too many EVs charge their batteries simultaneously could create a new peak demand.

Coordinated Charging Sortomme et al. [2011] propose a centralized charging scheduling to minimize the impact of EVs charging on distribution networks. More specifically, they aim to maximize load factor, minimize load variance, and minimize losses in the network. While centralized solution can guarantee optimal solutions, it is typically computationally expensive. To this end, Vandael et al. [2013] propose a solution that is a mixed between centralized and decentralized scheduling. First, the charging constraints of each EV are aggregated. Second, an optimization is performed on top of the aggregated constraints. Although this step can be viewed as a centralized optimization, it is still faster than optimizing the schedule of all EVs individually. Next, based on the optimized aggregated schedule and incentive signals sent by the aggregator, each EV computes its own schedule (hence, decentralized). Experiments suggest that this approach takes constant time with the increasing number of EVs, and scales linearly with the scheduling horizon length (while the fully centralized optimization could take polynomial time with order 3 and 5, respectively).

Ma et al. [2013] propose a decentralized mechanism to coordinate EV charging so as to minimize electricity generation cost. Similar to the case of energy consumption scheduling in Section 2.2.1, the cost function is assumed to be strictly convex and increasing. The mechanism is performed iteratively, where each EV updates its schedule slightly towards its optimal schedule at each iteration.[14] The aggregated demand is then updated and broadcasted to all

[14]Each EV updates its schedule only slightly towards its optimal schedule to avoid the herding effect, and thus

EVs. The iteration continues until no EV changes its schedule. Rather than computing the charging/discharging schedule well in advance, Ardakanian et al. [2013] propose a mechanism to continuously adapt the charging rate according to grid conditions.[15] More specifically, they propose a distributed control algorithm so that each EV can independently update its charging rate based on congestion of its network. An EV sets its charging rate higher when its feeders/transformers are lightly loaded, and lower when its feeders are heavily loaded.

Vehicle-to-Grid The Vehicle-to-Grid (V2G) concept sees EVs as both, challenges and opportunities. While the market penetration of EVs dramatically increases electricity demand, their battery could be leveraged for balancing the grid by discharging power back into the grid when supply is short. For instance, EVs can be charged during off-peak hours (possibly coordinated so as to not create a new peak), and providing energy services to the grid (as distributed generations or storage devices) during peak hours.

Since the impact of an individual EV to the grid is almost negligile, Guille and Gross [2009] and Kamboj et al. [2011] propose an implementation of V2G concept based on aggregation or coalition to provide a meaningful impact to the grid. The role central to this coalition is an aggregator that collects EVs and deals with the utility company (or energy supplier, or grid operators) as a single entity representing the EVs. The aggregator obtains profit by selling the aggregated capacity and energy services (such as regulation up and down in the regulation market) the collection of EVs can provide. To attract EV owners to join the coalition, Guille and Gross suggest the aggregator to provide incentives, such as reduced price for charging,[16] battery supply and maintenance, and parking services, in exchange for plugging in their EVs at the aggregator's premise at a specific time. In contrast, Kamboj et al. propose the aggregator to pay the EV owners proportional to their contribution in the coalition. Additionally, it is also possible for the aggregator to act as a controllable loads, charging the EVs without harming the grid (see also the previous paragraph about coordinated charging).

ensure the convergence (cf. [Li et al., 2011, Vytelingum et al., 2010]).

[15] The idea to avoid overloading the grid by actively sensing grid conditions is similar to that of [Ganu et al., 2013].

[16] Since the aggregated demand of the collection of EVs is large, rather than obtaining electricity supply from retailers, the aggregator could also obtain the supply directly from the market.

Part I

Demand Analytics

3 Customer Segmentation and Knowledge Discovery

This chapter introduces a structured framework and a discriminative index that can be used to segment the consumption data along multiple contextual dimensions such as locations, communities, seasons, weather patterns, holidays, etc. The generated segments can enable various higher-level applications such as usage-specific tariff structures, theft detection, customer-specific demand response programs, etc. The framework is also able to track customers' behavioral changes, evaluate different temporal aggregations, and identify main characteristics which define a cluster.

The work for this chapter was carried out during the author's internship at IBM Research, India. A shorter version of this chapter has also appeared in Proceedings of the 2014 SIAM International Conference on Data Mining [see Wijaya et al., 2014a].

3.1 Introduction

Many electricity suppliers around the world are deploying smart meters to gather fine-grained spatiotemporal consumption data (see also Figure 3.1). These companies are interested in mining the collected data to extract deep insights such as *the set of customers to be selected for winter peak load reduction, the set of customers to be monitored for potential theft/anomaly, the set of customers who can be targeted for energy efficiency programs, etc* [Beyond Zero Emissions, 2011, Durand, 2011, Oracle, 2011]. These insights are necessary for multiple application sub-domains in the energy sector such as billing, energy audit, etc. For all such advanced applications, *customer segmentation* has been viewed as one of key requirements [Chen et al., 1997, Figueiredo et al., 2005, Kitayama et al., 2002, Moss et al., 2008, Ramos et al., 2007].

However, segmenting customers based on the smart meter data is challenging due to three reasons. First, the scale of smart meter data is humongous: high volume (data from millions of customers) and high velocity (meters can report data at the rate of once every minutes to once every 30 minutes). Second, since electricity consumption is influenced by internal (family size, work hours, economic status, etc) and external contextual factors (weather, holiday, day of week, etc), the meter data must be correlated with heterogenous data sources (weather sites, survey results etc) that provide data at different time granularity and with varying data quality. Third, for meaningful grouping of customers, the segmentation may need to be performed along

Figure 3.1: Smart metering projects map around the world [Harrison, 2013]. Retrieved: 7 January 2015.

disparate contextual dimensions.

Overview of Contributions To address these challenges, we propose a novel framework for customer segmentation. The key contributions of this chapter are:

- Design and implementation of a versatile framework for customer segmentation that tries to jointly derive 'meaning' from consumption data, context data and user surveys. Previous works have primarily targeted a specific problem (e.g. setting tariff [Flath et al., 2012, Räsänen and Kolehmainen, 2009], predicting customer characteristics [Albert and Rajagopal, 2013]) and do not consider this task in a holistic manner.
- Design of a temporal aggregation method that varies the level of aggregation based on application requirements and data quality.
- Design of a novel *clustering consistency index* to track the evolution of consumption behaviors (that helps spotting fraudulent activities such as thefts and tampering).
- Design of a novel *discriminative index* and survey mining approach to identify main customer characteristics that can be used to classify customers into well-demarcated clusters.

Although previous works have taken initial steps in deriving customer segmentation based on smart meter data (see also Section 2.1.1), they indeed primarily target specific challenges/applications and do not present a general-purpose framework. Moreover, none of the prior works have incorporated additional context-specific data for customer segmentation. More importantly, as we will explain in the next section, different features and algorithms that are employed in the prior work, can be expressed as part of the building blocks in our framework.

The rest of the chapter is organized as follows. In Section 3.2, we explain our general purpose customer segmentation framework. In Section 3.3, we discuss our clustering consistency index.

In Section 3.4, we outline our method to extract knowledge from survey data to obtain main characteristics of a cluster. In Section 3.5, we describe the dataset and experimental results, and conclude in Section 3.6.

3.2 Context-Based Customer Segmentation

In this work, we developed a context-based 'general purpose' customer segmentation framework which exhibits 5 design principles addressing bottom-up data federation challenges and top-down unifying solution requirements discussed earlier. This is a user-centric design which provides a certain freedom to the framework users to respecify the data, context, and feature spaces she is interested in based on the specific customer segmentation task at hand. The framework enables the user to accomplish a number of customer segmentation tasks such as segmenting the customers based on the consumption magnitude, variability, or trend etc.

3.2.1 Design Principles

We define the requirement specification based on 5 design principles which can be used individually or in combination by the framework user:

R1 (Customized Data Selection): to declare *a*) a period she is interested in, such as from June to August 2013 *b*) a subset of customers that satisfy certain criteria (like average daily consumption greater than 5 kWh) *c*) a specific time of day she is interested in, such as afternoon peak hours 12:00 to 16:00

R2 (Customized Temporal Aggregation): to declare the time granularity of customers' consumption profile used for segmentation, such as hourly, every 3 hours, daily, weekly, or monthly

R3 (Customized Context): to declare specific context such as summer, winter, weekend, January, Tuesday, temperature more than a certain threshold, etc.

R4 (Customized Features): to declare specific feature computation such as mean, standard deviation, coefficient of variation or median.

R5 (Customized Algorithm) *a*) if the user has knowledge about a specific clustering algorithm to be used (from a predefined set) *b*) if supported by the algorithm, the user should be able to declare the number of customer segments (or clusters) that she is interested in, or let the framework determine the best number of clusters (according to some cluster evaluation metrics).

3.2.2 The Framework

Now, we define our framework, as shown in Figure 3.2. This framework is based on the design principles discussed above and supports the operations for data selection, temporal aggregation, context filtering, feature vector generation and clustering algorithm selection. Let $\mathcal{D} = \{d_1, \ldots, d_n, d_{n+1}, \ldots, d_{n+m}\}$ be a set of sensor devices available, where:

- $\{d_1, \ldots, d_n\}$ is the set of sensors that is the main subject of our analytics tasks, or *target sensors*, and
- $\{d_{n+1}, \ldots, d_{n+m}\}$ is the set of additional *context sensors*,

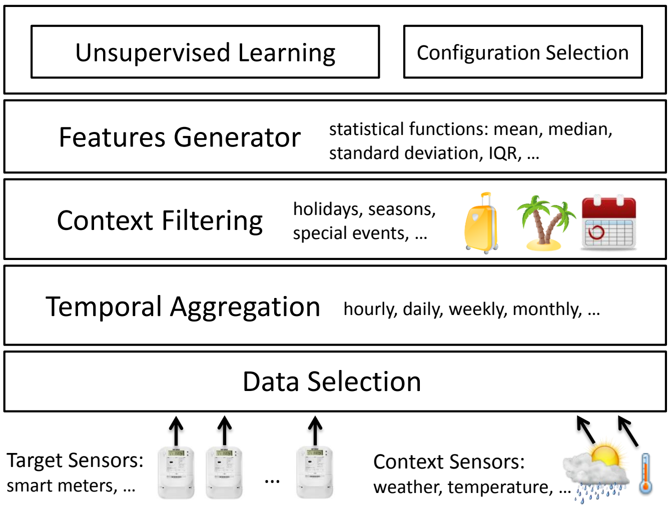


Figure 3.2: Architecture diagram of the framework.

In our case, smart meter is an example of a target sensor, whereas temperature, motion, or sound sensors are examples of context sensors.

For a vector $\mathbf{V}$, we write $\mathbf{V}(i)$ to address the i-th element of $\mathbf{V}$. Let $\gamma^{\star}(\dot{})$ represents application of a specific design principle $\star$ (from R1 to R5) to the input set.

Definition 3.1 (Measurements). *We define a* measurement-tuple *as $s = (t_s, \mathbf{V}_s)$, where:*

- *t_s is a* timestamp,
- *$\mathbf{V}_s \in \mathbb{R}^{n+m}$ is a vector of* sensor values, *i.e., $\mathbf{V}_s(i)$ is the value of d_i at time t_s.*

A time series of measurement *is defined as $S = \{s_1, \ldots, s_{|S|}\}$, where $s \in S$ is a measurement-tuple and whenever $i < j$, we have $t_{s_i} < t_{s_j}, \forall 1 \leq i, j \leq |S|$.*

Data Selection

Let $\mathcal{X}$ be the set of customers, ts_{start}, ts_{end} be the starting and the ending timestamp, and td_{start}, td_{end} be the starting and ending *time of day* that we are interested in, as the selection criteria of **R1**. In addition, let $timeOfDay(t)$ denote the time of day of timestamp t, i.e., the hour, minute, second, and millisecond of t.

Let S_x denotes the time series of measurements from customer x's premise. For a set of customer $\mathcal{X}$, we define $\mathbf{S}_{\mathcal{X}} = \{S_x \mid x \in \mathcal{X}\}$. Let $\mathcal{X}^+ \supseteq \mathcal{X}$. Then, data selection over $\mathbf{S}_{\mathcal{X}^+}$ is $\gamma^{R1}(\mathbf{S}_{\mathcal{X}^+}, \mathcal{X}, ts_{start}, ts_{end}, td_{start}, td_{end}) = \{S'_x \mid x \in \mathcal{X}\}$, where $S'_x = \{s_i \mid s_i \in S_x, ts_{start} \leq t_i \leq ts_{end}, td_{start} \leq timeOfDay(t_i) \leq td_{end}\}$.

Temporal Aggregation

Let $T = [\underline{T}, \overline{T}]$ be a time interval, where $\underline{T}$ and $\overline{T}$ both are timestamps as the lower and upper bounds of the interval, respectively. Let $\mathcal{T} = \{T_1, \ldots, T_{|\mathcal{T}|}\}$ be a set of time intervals denoting the temporal aggregation that we are interested in **R2**. For a time series of measurements S, temporal aggregation by $\mathcal{T}$ over S is defined as $\gamma^{R_2}(S, \mathcal{T}) = \{\hat{s}_1, \ldots, \hat{s}_{|\mathcal{T}|}\}$, where $\hat{s}_i = (T_{\hat{s}_i}, \mathbf{V}_{\hat{s}_i})$ is an *aggregated measurement*, $T_{\hat{s}_i} = T_i$ and

$$\mathbf{V}_{\hat{s}_i} = \sum_{t_s \in T_i} \mathbf{V}_s, \forall 1 \leq i \leq |\mathcal{T}|. \tag{3.1}$$

Note that in the Eq. (3.1) above we aggregate by summing up the sensor values. Depending on the application scenario, other aggregation function such as taking the average, maximum, or minimum values can also be used.

Example 3.1. *For monthly aggregation over one year data (from January to December), we have* $|\mathcal{T}| = 12$*, where each* $T \in \mathcal{T}$ *is a one month time interval. Thus, aggregation by* $\mathcal{T}$ *over any time series* S *results in* $|\gamma^{R_2}(S, \mathcal{T})| = 12$*, where each element, i.e., an aggregated measurement, contains the aggregation of sensor values over one month.*

Context Filtering

We define two context types that can be defined by the user with respect to **R3**, namely *calendar context*, and *measured context*. Calendar context is defined on timestamps, such as summer, January, weekday, or weekend. Measured context is defined on sensor values, such as temperature between 30 and 35 degree, humidity between 50% and 60%.

Definition 3.2 (Calendar Context)**.** *We define a* calendar context u *as a function* $f_u : t \to \{0, 1\}$*, where* t *is a timestamp. We have* $f_u(t) = 1$*, if* t *belongs to context* u*, and* 0 *otherwise. Let* U *be a set of calendar contexts, a time interval* $T = [\underline{T}, \overline{T}]$ satisfies U *iff* $f_u(\underline{T}) = 1$ *and* $f_u(\overline{T}) = 1, \forall u \in U$.

Example 3.2. *Let* $U = \{summer, weekend\}$ *be the set of calendar contexts that we are interested in. We have:*

- $f_{summer} : t \to \{0, 1\}$*, which return* 1 *if* t *is in summer, and* 0 *otherwise,*
- $f_{weekend} : t \to \{0, 1\}$*, which return* 1 *if* t *is on weekend days, and* 0 *otherwise.*

That is, time intervals which satisfies U *are intervals whose lower and upper bounds are both in summer and on weekend days.*

Definition 3.3 (Measured Context)**.** *We define a* measured context *as a tuple* $w = (\delta_w, \mathcal{X}_w)$ *where* $\delta_w \in \{n+1, \ldots, n+m\}$ *is a sensor index,* $d_{\delta_w} \in D$ *is the context sensor, and* $\mathcal{X}_w$ *is the accepted interval of the values of context sensor* d_{δ_w}*. A sensor values* $\mathbf{V} \in \mathbb{R}^{n+m}$ satisfies *a set of measured context* W *iff* $\mathbf{V}(\delta_w) \in \mathcal{X}_w, \forall w \in W$,

Definition 3.4 (Context Filtering)**.** *Let* $\hat{S}$ *be a time series of aggregated measurements,* U *be a set of calendar contexts, and* W *be a set of measured contexts. Context filtering over* $\hat{S}$ *by* U *and* W *is defined as* $\gamma^{R3}(\hat{S}, U, W) = \{\hat{s} \mid \hat{s} \in \hat{S},\ T_{\hat{s}} \text{ satisfies } U, \text{ and } \mathbf{V}_{\hat{s}} \text{ satisfies } W\}$.

Feature Vector Generation

In the context of energy consumption sensing, or environmental sensing in general, measurements from different time of day can be very different and hence, it is considered as an important feature for various data mining task. We take that knowledge into account by allowing the features to be built around different time intervals.

Let Γ be a set of time interval sets, where each $\mathcal{T}_i \in \Gamma$ is a time interval set. Let $\phi : 2^{\mathbb{R}} \to \mathbb{R}$ be a feature builder function (or feature function) with respect to **R4**, which typically is a statistical function, such as mean, median, standard deviation or inter-quartile range. And, let $\hat{S}$ be time series of aggregated measurements, where $\hat{s} = (T_{\hat{s}}, \mathbf{V}_{\hat{s}}), \forall \hat{s} \in \hat{S}$. A feature vectors computed from $\hat{S}$ using ϕ over Γ is $\gamma^{R4}(\hat{S}, \phi, \Gamma) = \mathbf{F} \in \mathbb{R}^{|\Gamma| \cdot n}$, where:

$$\mathbf{F}(i \cdot n + j) = \phi\big(\{\mathbf{V}_{\hat{s}}(j) \mid T_{\hat{s}} \in \mathcal{T}_i, \hat{s} \in \hat{S}\}\big), \qquad i = 0, \ldots, |\Gamma| - 1,\ j = 1, \ldots, n. \quad (3.2)$$

The $(i \cdot n + j)$-th element of feature vector $\mathbf{F}$ is computed using function ϕ over the set of aggregated sensor values $\mathbf{V}_{\hat{s}}$ that belong the same time interval set $\mathcal{T}_i$, and target sensor $d_j \in \{d_1, \ldots d_n\}$.

Example 3.3. *We give an example of hourly features generation. Let us assume that for the data selection, the user is interested in the data of year 2010. For the temporal aggregation, she is interested in hourly temporal aggregation. And, to simplify our example, let us assume that she is not interested in any context filtering, i.e.,* $U = \{\}$ *and* $W = \{\}$. *Furthermore, smart meter is our only target sensor, i.e.,* $n = 1$. *Assume that we have a time series of aggregated measurement,* $\hat{S}$, *for the whole year of 2010, where each* $\hat{s} \in \hat{S}$ *is an (hourly) aggregated measurement for each hour in 2010. Let* $\mathcal{T} = \{T_{\hat{s}} \mid \hat{s} \in \hat{S}\}$ *be a set of all time intervals in* $\hat{S}$, *thus we have* $|\mathcal{T}| = 365 \cdot 24 = 8760$. *Furthermore, let* $\Gamma = \{T_1, \ldots, T_{24}\}$ *be a set of time interval set, where each time interval set* $T_i \subset \mathcal{T}$ *contains the time intervals which accounts only for hour* i. *Let* ϕ *be a function that calculates mean. Hence, the result of* $\gamma^{R4}(\hat{S}, \phi, \Gamma)$ *is* $\mathbf{F} \in \mathbb{R}^{24}$, *where each* $\mathbf{F}(i)$ *is the mean of hourly consumption of hour* i *throughout the year 2010.*

Generation from a set of context sets and feature functions There could be a case where we would like to have a feature vector which is a combined result of applying **R3** using a set of context sets and **R4** using a set of feature functions. For example, instead of using mean as the only feature function, we might want to use both mean and median to have a more robust segmentation. Let $\hat{S}$ be the aggregated measurement satisfying **R1** and **R2**, Θ be the set of context sets, and Φ be the set of feature functions. In addition, let Γ be the set of time interval sets to build the features. For each context set $(U_i, W_i) \in \Theta$ and feature function $\phi_j \in \Phi$, we compute $\mathbf{F_{ij}} = \gamma^{R4}(\gamma^{R3}(\hat{S}, U_i, W_i), \phi_j, \Gamma)$. Finally, we append $\mathbf{F_{ij}}$, one after the other to form the combined feature vector, where $1 \le i \le |\Theta|$ and $1 \le j \le \Phi$.

Clustering Algorithm Application

Given the expert knowledge from the user to apply a specific algorithm, $A \in \mathcal{A}$, and its parameter setting ψ, then our framework should be able to apply it to the customers' feature vector (**R5** principle).

Let $\mathcal{X}$ be a set of customers, and $\mathbf{F}_{\mathcal{X}} = \{\mathbf{F}_x \mid x \in \mathcal{X}\}$ be a set of feature vectors of all customers in $\mathcal{X}$. Then, the application of clustering algorithm A with parameter setting ψ over $\mathbf{F}_{\mathcal{X}}$ results in a set of clusters (or *cluster configuration*, or *configuration*), i.e., $\gamma^{R5}(A, \psi, \mathbf{F}_{\mathcal{X}}) = \{c_1, \ldots, c_k\}$, where each cluster $c_j \subseteq \mathcal{X}$, $\forall 1 \le j \le k$, is a set of customers.

Automatic Cluster Configuration Selection Given different parameter settings, we are often uncertain which parameter setting delivers us the best cluster configuration (according to some cluster evaluation metrics). This holds even if the setting is simple and easy to understand, such as the number of clusters to be created (in case of using kMeans algorithm). For instance, we are often uncertain in choosing the value of k, the number of clusters. This motivates us to include an automatic selection of cluster configuration in our framework. Our selection mechanism is similar to the mechanism in [Martínez-Álvarez et al., 2011, Shen et al., 2013], i.e., we attempt to select compact and well separated clusters.

Given a clustering algorithm A, a set of parameter settings $\Psi = \{\psi_1, \ldots, \psi_{|\Psi|}\}$, and customers' feature vector $\mathbf{F}_{\mathcal{X}}$, we can have a set of cluster configuration $\mathcal{C} = \{C_i \mid \gamma^{R5}(A, \psi_i, \mathbf{F}_{\mathcal{X}}) = C_i\}$. Thus, our task is to select the best cluster configuration $C^* \in \mathcal{C}$. In order to determine C^*, we use three cluster evaluation metrics: Silhouette index [Rousseeuw, 1987], Dunn index [Dunn, 1973], and Davies-Bouldin index [Davies and Bouldin, 1979]. Although these indices provide us a way to compare a cluster configuration from one to another, there are some differences. Below, we give a brief description about these indices. Let x be a customer, C be a cluster configuration (set of clusters), and $C(x) \in C$ be the cluster of x. In addition, let $dist(x, x')$ be the distance between two customers x and x'.

The Silhouette index This index determines how well an object is clustered, based on the difference in the dissimilarity of the object to its cluster and to the other clusters. Let $dist(x, c)$ be the average distance between x and all customers in c, i.e.,

$$dist(x, c) = \frac{1}{|c|} \sum_{x' \in c} dist(x, x').$$

Furthermore, let $a(x)$ be the average dissimilarity of customer x to all other fellow cluster members in $C(x)$, i.e,

$$a(x) = \frac{1}{|C(x)| - 1} \sum_{\substack{x' \in C(x) \\ x' \neq x}} dist(x, x').$$

Assuming that $dist(x, x) = 0$, we rewrite the equation above into:

$$a(x) = \frac{dist(x, C(x))}{|C(x)| - 1}.$$

Let $b(x)$ be the minimum average dissimilarity between x and other clusters, i.e.,

$$b(x) = \min_{c \neq C(x)} \frac{dist(x, c)}{|c|}.$$

Then, we define the Silhouette value of x as:

$$silh(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}}$$

The Silhouette index of a cluster configuration is the average of the Silhouette index of all customers (in the configuration):

$$silh(C) = \frac{1}{|C|} \sum_{c \in C} \left(\frac{1}{|c|} \sum_{x \in c} silh(x) \right)$$

Silhouette index range from -1 to +1. The closer it is to 1, the better.

The Dunn index This index seeks the largest inter-cluster distance and the lowest intra-cluster distance. The Dunn index is computed based on the ratio between the minimum inter-cluster distance and the maximum intra-cluster distance. Let us define the inter-cluster distance between two clusters, c_1 and c_2, as the minimum distance between any two points in c_1 and c_2, i.e.,

$$\mathit{interdst}(c_1, c_2) = \min_{\substack{x_1 \in c_1 \\ x_2 \in c_2}} \mathit{dist}(x_1, x_2),$$

In addition, we define the intra-cluster distance (or *diameter*) of a cluster c, as the maximum distance between any two points in c, i.e.,

$$\mathit{dia}(c) = \max_{x_1, x_2 \in c} \mathit{dist}(x_1, x_2).$$

Then, we define the Dunn index of a configuration C as:

$$\mathit{dunn}(C) = \frac{\min\limits_{\substack{c_1, c_2 \in C \\ c_1 \neq c_2}} \mathit{interdst}(c_1, c_2)}{\max\limits_{c \in C} \mathit{dia}(c)}.$$

The larger the Dunn index, the better.

The Davies-Bouldin index This index is similar to the Dunn index, i.e., it aims to indentify a cluster configuration which has the largest inter-cluster distance and the lowest intra-cluster distance. The Davies-Bouldin index is computed based on the sum of diameter between two clusters divided by their inter-cluster distance:

$$\mathit{daviesBouldin}(C) = \frac{1}{|C|} \sum_{c_1 \in C} \max_{\substack{c_2 \in C \\ c_2 \neq c_1}} \frac{\mathit{dia}(c_1) + \mathit{dia}(c_2)}{\mathit{interdst}(c_1, c_2)}$$

In this case, we define the intra-cluster distance of a cluster c as the average distance of the cluster members to its centroid, i.e.,

$$\mathit{dia}(c) = \frac{1}{|c|} \sum_{x \in c} \mathit{dist}(x, \zeta^c),$$

where ζ^c is the centroid of cluster c. We define the inter-cluster distance to be similar with the one used for computing the Dunn index. Note that we define the Davies-Bouldin index here a little bit different compared to its original version [Davies and Bouldin, 1979]. However, as long as *dist* is a proper distance metric, our definition satisfies Definition 1 to 5 in [Davies and Bouldin, 1979]. The lower the Davies-Bouldin index, the better.

We next perform a majority voting to the best configuration identified by each metrics. Algorithm 3.1 describes the selection mechanism in more details. Functions $\mathit{sortSilhouette}(\mathcal{C})$, $\mathit{sortDunn}(\mathcal{C})$, and $\mathit{sortDaviesBouldin}(\mathcal{C})$ compute Silhouette, Dunn, and Davies-Bouldin index respectively for each configuration in $\mathcal{C}$, and return an ordered list of the cluster configurations, sorted by the configuration quality in descending order (that is, sorted in decreasing order for Silhouette and Dunn indices, and in increasing order for Davies-Bouldin index). Let $\mathit{count}(L, e)$ be the count of element e in the list L. Function $\mathit{mostFrequent}(L)$ returns an element e^* in L, where $\mathit{count}(L, e^*) > \mathit{count}(L, e)$ for all $e \neq e^*$ in L. In other words, $\mathit{mostFrequent}(L)$ returns the element with the largest count, e^*, and there is no other element in L which has the same count as e^*. If there is no such element, this function returns **null**.

Algorithm 3.1: Automatic Cluster Configuration Selection

Input: a set of cluster configuration $\mathcal{C} = \{C_1, \ldots, C_{|\mathcal{C}|}\}$

Output: the best configuration $C^* \in \mathcal{C}$

$silhList \leftarrow sortSilhouette(\mathcal{C})$
$dunnList \leftarrow sortDunn(\mathcal{C})$
$davbList \leftarrow sortDaviesBouldin(\mathcal{C})$
$countList \leftarrow [\,]$
$\mathbf{C}^* \leftarrow$ **null**
$i \leftarrow 1$
repeat
  $countList.add(silhList[i])$
  $countList.add(dunnList[i])$
  $countList.add(davbList[i])$
  $C^* \leftarrow mostFrequent(countList)$
  $i \leftarrow i + 1$
until $(i > |\mathcal{C}|) \vee (\mathbf{C}^* = \mathbf{null})$
return $\mathbf{C}^*$

3.3 Clustering Consistency

This section answers two important questions. First, *is there any customer who changes their behavior over time?* For example, we would like to know whether there is a customer who is in the low consumption cluster in January, but she changes to the medium/high consumption cluster in February. This insight is important for devising personalized feedback to the customer. Second, *how different is one cluster configuration to another?* For example, how different is the cluster configuration using January data compared to using February data, or March, April, etc. Answering this question gives insight to the utility company on the key contexts to consider when developing policies, such as differential pricing or demand response signal. In the sequel, we use the term *individual* and *customer* interchangeably.

3.3.1 Individual to Cluster Consistency

Given two cluster configurations, we develop a measure to indicate whether an individual has the same cluster membership on both of them. Because cluster configuration is invariant to the cluster labels, we require the measure to also be invariant to the label sets. Thus, the idea is to define an *individual to cluster consistency* index (*i2c*) which computes how consistent are the fellow cluster members of an individual on the two configurations.

Let C be a cluster configuration. We write $C(x)$ to denote the cluster of x, i.e., the cluster $c \in C$ where $x \in c$. We define an individual to cluster consistency of customer $x \in \mathcal{X}$ on two cluster configurations C_1 and C_2 as:

$$i2c(x, C_1, C_2) = \frac{|C_1(x) \cap C_2(x)| + |(\mathcal{X} \setminus C_1(x)) \cap (\mathcal{X} \setminus C_2(x))| - 1}{|\mathcal{X}| - 1}. \tag{3.3}$$

Intuitively, if we denote the set of customers in $C(x)$ as the *friends* of x, and the others as *non-friends*, then *i2c* measure the number of friends in C_1 who also friends of x in C_2, and the number of non-friends in C_1 who also non-friends in C_2 normalized by the number of all customers excluding x. The value of *i2c* ranges between 0 and 1. The closer it is to 1, the more consistent is x's cluster membership in C_1 and C_2.

3.3.2 Distance Rank

Given individuals who change their clusters, one might interested more to the ones located closer the centroid of their new clusters. Thus, we define an additional measure, *distance rank*, to denote the ranking of individual's distance to its cluster representative (such as centroid) compared to the other cluster members. We use distance rank instead of actual distance measure because it is invariant to cluster size. Thus, it can be used for comparison across different clusters.

Let $C(x)$ be the cluster of x in configuration C, and $\zeta^{C(x)}$ be the cluster representative of $C(x)$. In addition, let $dist(x, \zeta^{C(x)})$ be the distance of x to its cluster representative. For an individual x, we define its distance rank as:

$$dr(x, C(x)) = \frac{|\{x' \mid dist(x, \zeta^{C(x)}) < dist(x', \zeta^{C(x)}), x' \in C(x)\}|}{|C(x)|}. \tag{3.4}$$

Distance rank ranges between 0 and 1. The higher the value, the closer the individual to its cluster representative.

3.3.3 Cluster Configuration Consistency

In order to investigate community behavioral changes over different contexts, we can measure the difference of the resulting cluster configuration over those contexts. Let $\mathcal{X}$ be a set of customers, and C_1 and C_2 be two cluster configurations over $\mathcal{X}$. We compute the difference between C_1 and C_2, i.e., *cluster configuration consistency* index, as the average of *i2c* of their individuals:

$$ccc(C_1, C_2, \mathcal{X}) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} i2c(x, C_1, C_2). \tag{3.5}$$

The *ccc* index ranges between 0 and 1. The higher the *ccc* between C_1 and C_2, the more similar they are.

3.4 Knowledge Extraction from Survey Data

Customer segments can be useful for implementing different policies, such as: targeted demand responses, more personalized energy feedback, or differential pricing. However, having customer segmentation alone is not enough. We also need to understand *what are the characteristics that constitute a customer cluster*? Only by developing this understanding, we can develop an effective and efficient policies which tailored better for our customers.

In their work, Albert and Rajagopal correlated customers' consumption profile with demographics and appliance usages [Albert and Rajagopal, 2013]. However, they did not attempt to identify relevant characteristics for clustering customers. Instead, they started with a predefined set of characteristics and determined whether those can be predicted from customer's consumption profile. This approach is rather similar with [Fusco et al., 2012] where the authors used the same dataset as ours to predict household demographics from consumption profiles. Regarding customer behavior analysis, Pedersen [2008] presented a psychographic customer segmentation based on how customers feel, think, and act. However, their segmentation is based solely on survey data about customers' behavior and attitude toward electricity and energy conservation, and did not involve processing of any consumption data.

Customer characteristics can be of form demographic profiles, house types, appliance usages, or living styles. We can obtain these through survey/questionnaire, using questions, such as: What best describes the people you live with (single/adults/adults with children)? Do you have a dishwasher (yes/no)? What is the approximate floor area of your home?[1] In this section, we focus on mining customers characteristics, which are the discriminative, i.e., characteristics which make a cluster different from the others. We model customer characteristic as a pair of question and answer. Next, we describe how to compute *discriminative index*.

3.4.1 Discriminative Index

We define a measure to express how discriminative a question-answer pair is in distinguishing a cluster from the others. Let $\mathcal{X}$ be a set of customers, and C be a cluster configuration over $\mathcal{X}$. For a cluster $c \in C$, we denote $\neg c$ as all individuals who are not in c, i.e., $\neg c = \{x' \mid x' \in \mathcal{X} \setminus c\}$. In addition, let q be a question, $ans(q)$ be the set of possible answers to q, $ans_x(q) \in ans(q)$ be the answer of customer x to question q, and $N_{c,q}$ be the set of customers in cluster c who respond to question q.

We define $Z_c(q, a)$ as the fraction of individuals in cluster c who answer a to question q, i.e.,

$$Z_c(q, a) = \frac{|\{x \mid ans_x(q) = a,\ x \in c\}|}{|N_{c,q}|}, \tag{3.6}$$

where $a \in ans(q)$. Then, *discriminative index* of question q and answer a to cluster c is defined as:

$$DI_c(q, a) = \frac{Z_c(q, a) - Z_{\neg c}(q, a)}{\max\big(Z_c(q, a), Z_{\neg c}(q, a)\big)}. \tag{3.7}$$

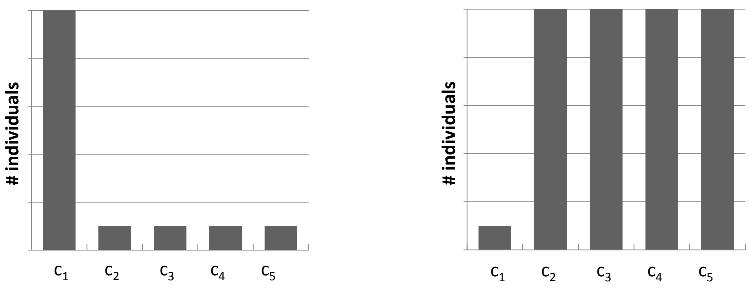
Discriminative index ranges between -1 and $+1$. It is discriminative positive (or negative) if it is positive (or negative). Both discriminative positive and negative explain how a cluster differs from the others. Discriminative index close to $+1$ means that most of the individuals in cluster c answer a to question q, whereas individuals in other clusters do not. In contrast, discriminative index close to -1 means that most of individuals in other clusters answer a to question q, whereas individuals in cluster c do not. Discriminative index close to 0, means that answer a to question q does not differentiate cluster c from the others, i.e., it has little or no discriminative power for cluster c.

3.4.2 Dealing with Ordinal and Quantitative Data

In a survey, there are some answers which are ordinals or quantitatives. For example, in our survey data, answers to the question whether a customer would like to do more to reduce electricity usage, are ordinals, i.e., five criteria from strongly agree to strongly disagree. Another example: answers to a question of the approximate floor area of the house, are quantitative, i.e, the number which represent the floor area.

In the previous section, we determine whether a specific pair of question and answer is a key characteristic of a cluster. However in ordinal and quantitative answers, we are interested on insights more than that. For example, instead of "most of the customers in cluster c have X sq ft. floor area ", we are more interested on more general insight, if any, such as: "most of the customers in cluster c have *less than* X sq ft. floor area", or "*between* X *and* Y", or "*greater than* X". One way to do this is by introducing some splitting points which divide the answers into groups, or ranges. But then, it leads us into a combinatorial problem, such as: how many points do we need for the best splitting, and where should we put the splitting points.

[1] These example questions are taken from our dataset (explained in Section 3.5).



(a) Feature f_1 (b) Feature f_2

Figure 3.3: Feature f_1 is discriminative positive for cluster c_1, whereas f_2 is discriminative negative for c_1. While entropy measure is able to recognize only discriminative positive features, our *discriminative index* is able to recognize both, discriminative positive and negative features.

Let q be a question, and $ans(q)$ be a set of customers' answers to q. To solve this problem, we sort the answers in ascending order and put them into a list. We add $-\infty$ and $+\infty$ to the beginning and the end of the list. Let $l = |ans(q)| + 2$ be the length of the list. We take all possible n-grams from the list, where n varies from $l - 1$ to 1. Next, we create a set of ranges $R_{ans(q)}$ by taking the first and the last element of each n-grams as the lower and upper bounds (inclusive). Finally, we remove ranges $[-\infty, -\infty]$ and $[\infty, \infty]$ from $R_{ans}(q)$. This takes polynomial time on the size of $ans(q)$.

Example 3.4. *Let $ans(q) = \{1, 2, 5\}$. The set of ranges created from all possible n-grams from length 4 to 1, without $[-\infty, -\infty]$ and $[\infty, \infty]$, is $R_{ans(q)} = \{\ [-\infty, 5],\ [1, \infty],\ [-\infty, 2],\ [1, 5],\ [2, \infty],\ [-\infty, 1],\ [1, 2],\ [2, 5],\ [5, \infty],\ [1, 1],\ [2, 2],\ [5, 5]\ \}$.*

Then, for ordinal and quantitative questions/answers, instead of computing discriminative indices based on $a \in ans(q)$, we now compute them based on $a^r \in R_{ans(q)}$:

$$Z_c(q, a^r) = \frac{|\{x \mid ans_x(q) \in a^r,\ x \in c\}|}{|N_{c,q}|}, \tag{3.8}$$

$$DI_c(q, a^r) = \frac{Z_c(q, a^r) - Z_{\neg c}(q, a^r)}{\max\big(Z_c(q, a^r), Z_{\neg c}(q, a^r)\big)}. \tag{3.9}$$

3.4.3 An Alternative to Discriminative Index

Entropy can also be used as an alternative to our *discriminative index* to determine whether a certain customer characteristics is discriminative or not, using the same idea as in the decision tree learning. However, there is a subtle difference.

Using entropy, a feature is said to be discriminative for a particular class (or cluster, in our case) when it has low entropy. In Figure 3.3, f_1 has low entropy, and hence it is discriminative. That is, f_1 is an appropriate feature to distinguish cluster c_1 from others. Moreover, f_1 as an example of what we called as a *discriminative positive* feature. Feature f_2 in Figure 3.3, has high entropy. Thus, according to the entropy measure, f_2 is not discriminative. However, we can see that f_2 is actually also a discriminative feature, in the sense that it characterizes an individual which does not belong to c_1 (it might belong to any other clusters). Feature f_2 is an example of what we called as a *discriminative negative* feature.

While entropy is useful measure to recognize discriminative positive feature, it does not recognize discriminative negative feature. Our *discriminative index*, on the other hand, is able to distinguish both, discriminative positive and negative features.

3.5 Experimental Evaluations

In this section, we describe our experiment details.

3.5.1 Dataset

We use the detailed data underlying electricity consumption behaviour provided in anonymized format by the Commission for Energy Regulation (CER) in Ireland.[2] This dataset is the result of the Electricity Customer Behaviour Trials (CBTs) [The Commission for Energy Regulation (CER) and the ISSDA, 2012], which started in July 2009 and ended in December 2010 with over 5,000 Irish homes and businesses participating. The participants in the trials had an electricity smart meter installed in their homes/premises, which collected energy consumption measurements (in kWh) every half hour. The objective of the trial was evaluating the impact that different Time-Of-Use (TOU) tariffs have on the consumption behaviour.

Although the CER has carefully cleaned the data (e.g., multiple imputation for the missing values—see [The Commission for Energy Regulation (CER) and the ISSDA, 2012], Appendix 2), there are still a small number of missing values found in the dataset. In this work, unless stated otherwise, we choose customers who have no missing values in their measurements. Furthermore, to avoid bias due to the TOU tariffs, we consider only the residential households in the control group of the trial, i.e.,, those customers with a flat rate that did not change their consumption behavior in response to a TOU tariff. This results in the selection of 782 customers. The measurements are aggregated into hourly timeslots.

In addition, the dataset also contains survey results, which includes customers' demographics (occupation, family type, etc.), house information (ownership, age, floor area, etc.), and appliance usages (dishwasher, TV, water pump etc.).[3]

3.5.2 Customer Segmentation

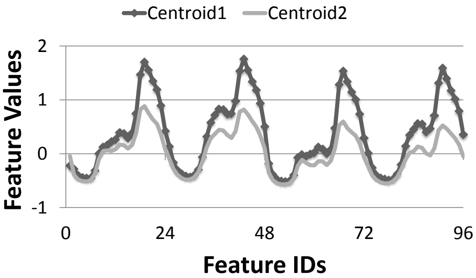
First, we demonstrate the result of our selection mechanism (described in Section 3.2.2). Using the CER dataset, smart meter is the only target sensor, $n = 1$. Although there is no context sensor, i.e., $m = 0$, we showcase several calendar context. Second, we show the result of our framework to accomplish different customer segmentation tasks, i.e., customer segmentation by consumption trends, absolute consumption, and consumption variability. We also show that applying the same task on different contexts yield different results. We visualize each cluster by its centroids.

Automatic Cluster Configuration Selection

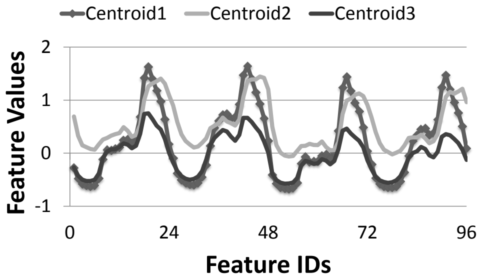
Our automatic selection mechanism aims to select compact clusters which far from each other (well separated). We perform a customer segmentation task based on their consumption trends. We select all data (R1). We use hourly temporal aggregation (R2). Our features is a combined

[2] http://www.ucd.ie/issda/data/commissionforenergyregulationcer/

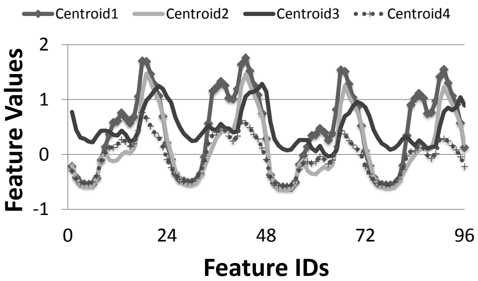
[3] http://www.ucd.ie/issda/static/documentation/cer/smartmeter/cer-residential-pre-trial-survey.pdf



(a) $k = 2$



(b) $k = 3$



(c) $k = 4$

Figure 3.4: Centroids of clusters using January data, and hourly temporal aggregation. For the features, we use normalized mean of weekday (ID 1-24), normalized mean of weekend (ID 25-48), normalized median of weekday (ID 49-72), and normalized median of weekend (ID 73-96) consumption. We use kMeans algorithm with $k = 2$, $k = 3$, and $k = 4$.

feature vector, from a set of context sets (R3) and feature functions (R4). We use calendar context only (without measured context), i.e., the set of context sets $\{(U, W)\}$ is $\big\{\big(\{$January, weekday$\}$, $\{\}\big)$, $\big(\{$January, weekend$\}$, $\{\}\big)\big\}$. To obtain the consumption trends, we use $\{$normalized mean, normalized median$\}$ as the set of feature functions. *Normalized* here means that we apply standard normalization on the measurement tuple, such that its mean is 0 and its standard deviation is 1. We use kMeans algorithm, for $k = 2, \ldots, 10$ (R5). As a consequence, we obtained 9 different cluster configurations.

Cluster configuration resulting from $k = 2$ is determined as the best configuration by our automatic selection mechanism. Figure 3.4 illustrates the result using $k = 2$, $k = 3$, and $k = 4$. We can see that centroids of clusters using $k = 2$ are better separated than the others. In addition, the configuration separates the customers who have high and low peak consumption in the evening.

Various Segmentation Tasks

In this section, we show the generality of our framework to accomplish different customer segmentation tasks. Furthermore, we show that applying customer segmentation in different contexts produce different results.

We perform customer segmentation by consumption trends, absolute consumption, and consumption variability. The setting are similar as in Section 3.5.2, except for the set of context sets and feature functions. We use the set of feature functions $\{$normalized mean, normalized median$\}$ for consumption trends, $\{$mean, median$\}$ for absolute consumption, and $\{$standard deviation, IQR} for consumption variability. We perform the task in three different contexts: January, July, and all months, and separate weekend consumption from weekdays. Thus, we use the set of context sets: {({January, weekday}, {}), ({January, weekend}, {})} for January, {({July, weekday}, {}), ({July, weekend}, {})} for July, and {({weekday},{}), ({weekend}, {})} for all months.

Figure 3.5 shows that for segmentation by consumption trends, we successfully divide the customers into high peak and low peak customers. For the next tasks, segmentation by absolute consumption and consumption variability, we are also able to produce clusters with high and low absolute consumption and consumption variability, respectively. Furthermore, comparing the results on January, July, and all months, shows that applying the segmentation tasks in different contexts yields different results. This validates our approach to perform context-based customer segmentation.

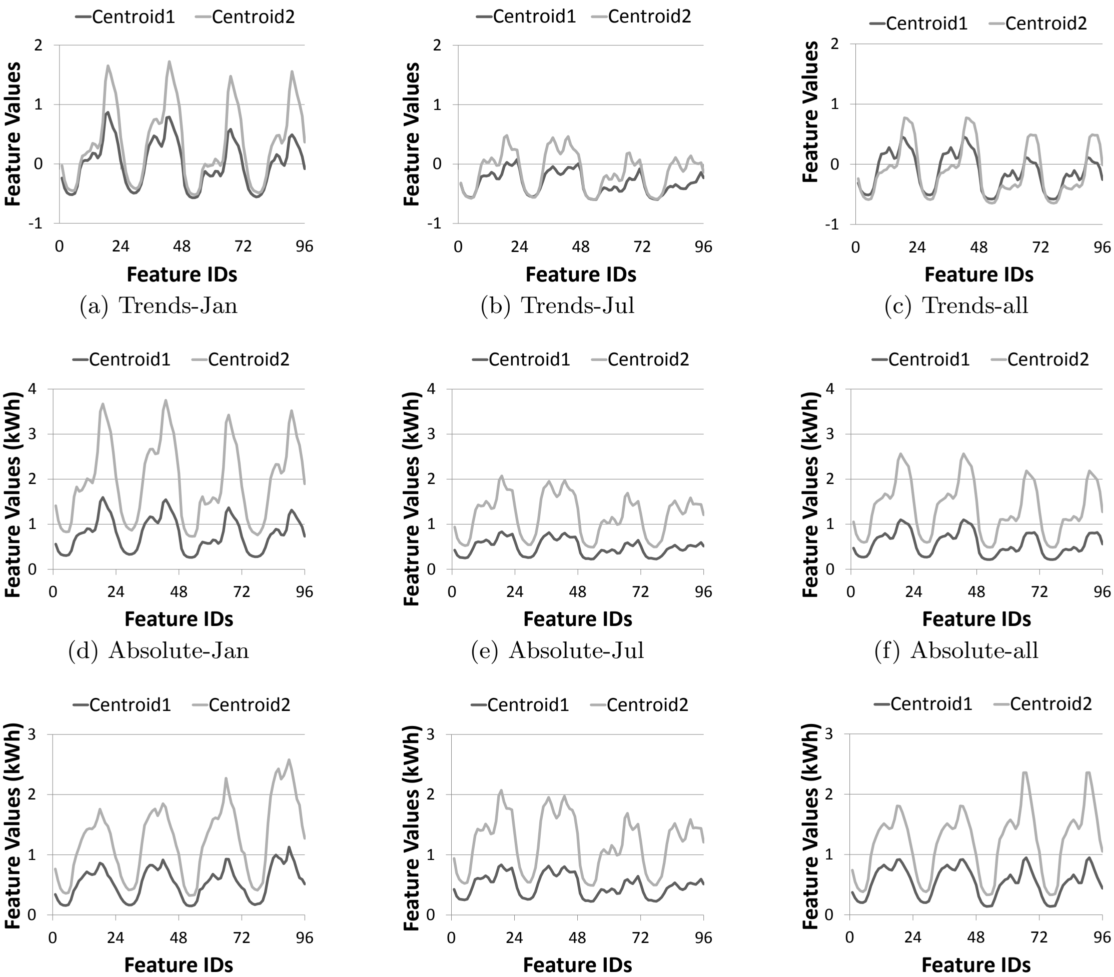
3.5.3 Clustering Consistency

Using our clustering consistency index (described in Section 3.3), we can quantify the difference between cluster configurations. Figure 3.6a shows the difference between cluster configurations, resulting from customer segmentation by absolute consumption for a specific month compared to 1, 3, and 6 months previously. We use kMeans algorithm, with k=2. This results in two customer segments: high and low consumption clusters.

The result shows that the consistency between clusters resulting from the segmentation of the current month and 1 month ago is higher than the consistency between clusters from the current month and 3 months or 6 months ago. Especially, the lowest consistency is between July 2010 and 6 months previously, January 2010 (cluster configuration consistency = 0.67). One of the reason is seasonal changes in the energy consumption behavior between summer and winter, i.e., July and January is in the middle of summer and and winter in Ireland, respectively. However the implication of our result could be bigger than that. It indicates that, there are a number of customers who behave differently (compared to their fellow cluster members), which in turn change their cluster memberships.

To elaborate this, in Figure 3.6b and 3.6c, we show the consumption profile (mean and median of weekday and weekend consumption) of the centroid of the low consumption cluster, and two individuals: ID 1301 and ID 7381, in January and July 2010. These two customers are in the low consumption clusters in January 2010. Both of them have the highest distance rank among individuals with low consistency index (ID 1301) and high consistency index (ID 7381) between January and July 2010. Customer ID 1301 changes her cluster membership from low consumption cluster in January 2010 to the high consumption cluster in July 2010 (we use centroid as the cluster representative). The typical consumption of the low consumption cluster in July 2010 is lower than January 2010, which shows the seasonal changes in the electricity consumption between winter and summer. Customer ID 7381, who stays in the low consumption cluster, lower her consumption level, in line with the behavior of her cluster. However, the consumption of customer ID 1301 in July 2010 is approximately the same as her consumption in January 2010. This causes her to change her cluster membership to the high consumption cluster in July 2010. Then, given this results, we can devise a personalized energy feedback for customer ID 1301 to lower her energy consumption (for example, using self-comparison). Another possible explanation might be that while other customers uses electric heating, customer ID 1301 does not—which also reveals another valuable information about the characteristics of the customer in question.

In addition, our cluster configuration consistency index is useful to quantify the difference between results obtained by various temporal aggregations. Since smart meter data has high



(a) Trends-Jan (b) Trends-Jul (c) Trends-all

(d) Absolute-Jan (e) Absolute-Jul (f) Absolute-all

(g) Variability-Jan (h) Variability-Jul (i) Variability-all

Figure 3.5: Customer segmentation on consumption trends (a)-(c), absolute consumption (d)-(f), and consumption variability (g)-(i), in different contexts: January, July, and all months. For trends, we used the same features as in Figure 3.4. Feature ID 1-24 and 49-72 are weekdays consumption, whereas 25-48 and 73-96 are weekend consumption. We use mean (ID 1-48) and median (ID 49-96) for absolute, and standard deviation and IQR for variability.

velocity and high volume, determining the right aggregation is imperative. We compare the segmentation results by absolute consumption, consumption variability, and trends, performed using different temporal aggregations, against hourly temporal aggregation.

Figure 3.6d shows that, for customer segmentation by absolute consumption, we have only a little difference between cluster configurations resulting from various temporal aggregations and hourly aggregation. This is a good news because storing and processing monthly aggregated data, for example, is far more desirable than hourly data. Unfortunately, for customer segmentation by consumption variability and trends, this is not the case. The consistency index decreases as the temporal aggregation become coarser, i.e., the coarser the temporal aggregation, the

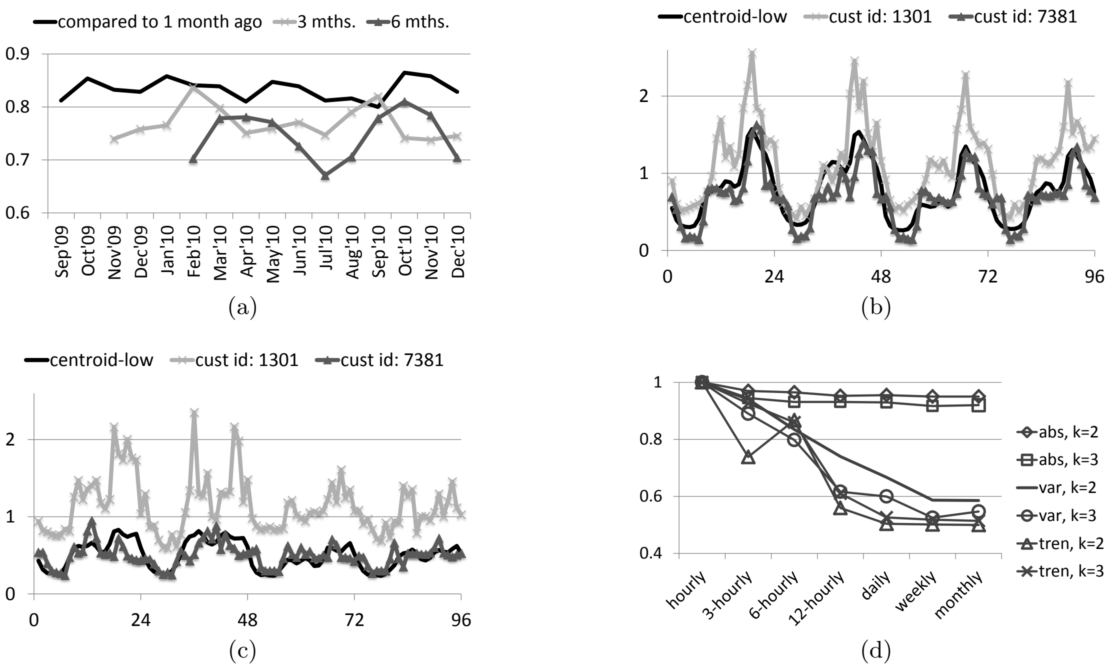


Figure 3.6: (a) cluster configuration consistency over time (monthly), consumption profile using (b) January 2010 data, (c) July 2010 data, and (d) cluster configuration consistency over different temporal aggregations.

Table 3.1: Customer characteristics based on their absolute consumption. A minus (-) sign denotes discriminative negative.

Cluster	Question	Answer	DI
low	family type	single	0.86
	floor area (sq ft.)	805-1073	0.86
	#bedrooms	≤ 2	0.85
medium	electric shower	(-) ≥ 20 mins	-0.76
	family type	(-) single	-0.61
	floor area (sq ft.)	2300-2750	0.56
high	#children	≥ 4	0.93
	family type	(-) single	-0.90
	floor area (sq ft.)	(-) ≤ 1200	-0.87

higher the difference with the hourly aggregation. This can be understood since as we move to coarser temporal aggregation we lose the variation in the consumption profiles, which is needed to distinguish different consumption variability or trends.

3.5.4 Knowledge Extraction from Survey Data

Our dataset contains not only smart meter measurement, but also customer survey data (questions/answers). From this survey data, using our discriminative index (described in Section 3.4) we are able to extract knowledge about main characteristics of a cluster. This step is imperative for applying the right business decision or policy to a specific customer segment.

Table 3.2: Customer characteristics based on their consumption variability. A minus (-) sign denotes discriminative negative.

Cluster	Question	Answer	DI
low	water pump	(-) 1-2hrs	-0.88
	family type	single	0.80
	washing machine	(-) $\geq$ 2-3 loads	-0.76
medium	electric shower	10-20 mins	0.59
	family type	(-) single	-0.55
	#children	(-) ≥ 3	-0.54
high	tumble dryer	$\geq$ 2 to 3 loads	0.90
	#children	≥ 4	0.88
	floor area (sq ft.)	≥ 2800	0.79

In Table 3.1 and 3.2, we show the top 3 characteristics of customer segments, formed by absolute consumption and consumption variability, respectively. We use the same features and setting as in Figure 3.5f and 3.5i, with k=3 (number of clusters). The characteristics (expressed by questions and answers) are ordered by the absolute value of their discriminative index (DI). Recall that DI < 0 denotes discriminative negative, i.e., the answer is associated more likely to other clusters.

Customer segments by absolute consumption is determined more by customer's demographics (such as family type, the number of children) and housing condition (floor area, the number of bedrooms). Low consumption cluster is dominated by single, whereas medium/high consumption clusters are dominated by non-single family, either adults only or adults with children. Floor area is also relevant, with low consumption clusters having the smaller area.[4]

There are more insights which are based on appliance usages in Table 3.2. It can be understood since consumption variability comes from intermittent appliance usages. Customers with low consumption variability use big appliances, such as water pump and washing machine, in a shorter duration.[5] Furthermore, customers with high consumption variability use tumble dryers for the longest duration.

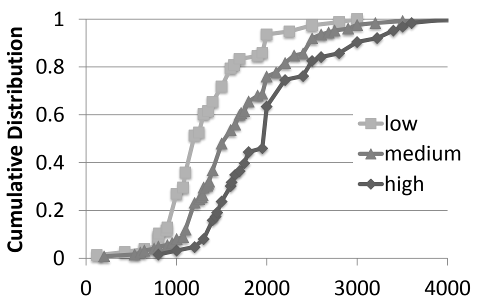
Floor Area vs. The Year the House Was Built Often opinion about house's energy consumption is built around its floor area or the year it was built.While Table 3.1 and 3.2 shows insights about the floor area, there is none about the year it was built. To investigate this further, we plot the cumulative distribution of customers' floor area (Figure 3.7a) and the year their houses was built (Figure 3.7b). Figure 3.7a shows that, indeed floor area is a relevant characteristics, i.e., we can distinguish clearly the cumulative distribution of the floor area among different clusters, where customers in the lower consumption cluster typically associated with smaller floor area. Figure 3.7b, however, shows that this is not the case with the year the houses was built, where the cumulative distribution for the three clusters are similar (coincide) to each other.

Appliance Ownership Compared to appliance usage, information about appliance ownership is simpler and cheaper to obtain. Using questionnaire is enough to obtain the information whether a customer own a certain appliance. Detailed appliance usage information, however, is more expensive to obtain because it involves sensor measurement.[6] Thus, knowing whether

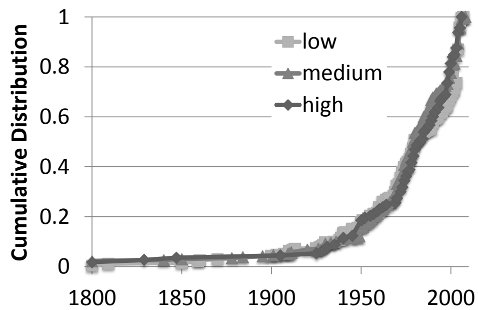
[4] Maximum customers' floor area is around 5000 sq ft.

[5] Water pump usages are ranges from < 30 minutes to > 2 hours. Washing machine usages are ranges from < 1 load to > 3 loads a day.

[6] Typical appliance usage, however, as in our dataset, can be obtained through questionnaire.



(a) Floor area (sq ft.)



(b) Year built

Figure 3.7: Cumulative distribution of (a) floor area, and (b) the year customers' houses was built. Customers are clustered by their absolute consumption: low, medium, and high.

ownership of a particular appliance determines customer's energy consumption is a valuable insight.

In our dataset, we have a set of question/answer whether a customer own these appliances:

- washing machine,
- tumble dryer,
- dishwasher,
- electric shower,
- electric cooker,
- stand alone freezer,
- water pump,
- immersion,
- TV less than 21 inch,
- TV greater than 21 inch,
- desktop computer,
- laptop computer, and
- games consoles.

In Table 3.3 and 3.4, we show customer characteristics which related to appliance ownership only. Both shows how discriminative is an ownership of a particular appliance for different clusters, based on absolute consumption and consumption variability. Let *support* be Z_c in case of discriminative positive and $Z_{\neg c}$ in case of discriminative negative. We show only characteristics with DI ≥ 0.6 (highly discriminative) and *support* ≥ 0.4 (highly evident).

Over all appliances, we found that, only the ownership of big (power consuming) appliances (dishwasher and tumble dryer), which are highly discriminative. That is, the owner of these

Table 3.3: Discriminative appliances' ownership for different clusters based on their absolute consumption. We show only for DI ≥ 0.60 and support ≥ 0.40. A minus (-) sign denotes discriminative negative.

#	Appliance	Cluster	Ownership	DI
1	dishwasher	high	(-) no	-0.76
2	games consoles	low	(-) yes	-0.70
3	tumble dryer	low	no	0.68
4	dishwasher	low	no	0.67
5	games consoles	high	yes	0.61

Table 3.4: Discriminative appliances' ownership for different clusters based on their consumption variability. We show only for DI ≥ 0.60 and support ≥ 0.40. A minus (-) sign denotes discriminative negative.

#	Appliance	Cluster	Ownership	DI
1	dishwasher	high	(-) no	-0.72
2	tumble dryer	high	(-) no	-0.72
3	tumble dryer	low	no	0.71
4	games consoles	low	(-) yes	-0.69
5	dishwasher	low	no	0.67
6	games consoles	high	yes	0.60

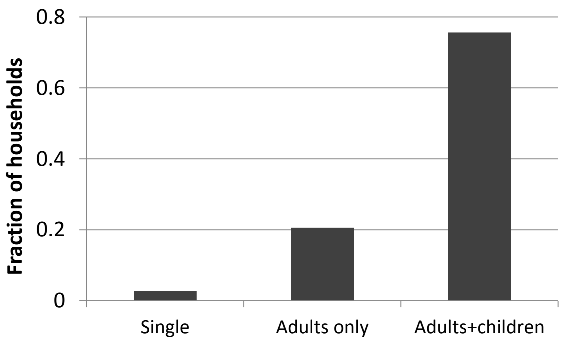


Figure 3.8: Fraction of households which own games consoles for different family types.

appliances are more likely to consume more energy and have higher consumption variability. The ownership of other appliances, which are not shown in Table 3.3 and 3.4, are less discriminative.[7]

The consistent presence of games consoles in both tables, however, is rather interesting since they are not big appliances (their power consumption is comparable to other electronic devices such as TV or desktop computer). We conjecture that the ownership of games consoles is highly correlated with family type, e.g., families with children are more likely to have games consoles at home compared to singles. Because family type is a highly discriminative characteristics for households' energy consumption behavior (see Table 3.1 and 3.2), then its correlation with games consoles ownership explains why games consoles ownership is also discriminative. Our conjecture is then confirmed in Figure 3.8, where it shows that, indeed, families with children are the most likely to own games consoles, followed by adults only families, and then by singles,

[7] However, their usage pattern might be highly discriminative (such as washing machine, electric shower, water pump – see Table 3.1 and 3.2).

who are the least likely to own games consoles.

3.6 Summary and Discussion

This chapter presents a generic customer segmentation framework that can be used to classify smart meter data into clusters using multiple distinguishing characteristics such as time of consumption, levels of consumption, associated contexts, etc. We also present a clustering consistency index, which can be used to track evolving consumption behaviors and to compare customer segments resulting from different temporal aggregations. Moreover, it is also similar to Rand index [Rand, 1971]. However, we generalize it further to determine whether an individual is likely to change cluster over time, i.e., individual to cluster consistency measure.

We evaluate the framework and index using real world smart meter data and survey results. Experiments show that customer segmentation results are different from one context to another. Moreover, different temporal aggregations have only a little effect on segmentation by absolute consumption. But, this does not hold for segmentation by consumption variability or trends. Furthermore, customer's floor area is relevant to her consumption. In addition, big appliances' usage patterns also play a role in customer's consumption variability.

4 Forecasting Residential Demand

One of the important tasks in various Smart Grid applications, from demand response to emergency management, is the short-term electricity load forecasting at different scales, from an individual household to a whole portfolio of customers. While a large body of the load forecasting literature has focused on large, industrial, or national demand, this chapter focuses on energy consumption of residential customers. More specifically, we quantitatively evaluate different machine learning methods for short-term (1 and 24 hour ahead) electricity load forecasting at the individual and aggregate level. Additionally, since energy consumption behavior may vary between households, we first build a feature universe, and then apply Correlation-based Feature Selection to select features relevant to each household. We also find that the improvement provided by the Cluster-based Aggregate Forecasting strategy depends not only on the number of clusters, but more importantly on the size of the customer base. Consequently, our finding provides a valuable insight to practitioners who wish to implement the strategy in the real world.

This chapter revises and extends [Humeau et al., 2013]. The preliminary version of this chapter can also be seen in [Wijaya et al., 2014b].

4.1 Introduction

The exploitation of renewable energy, the integration of distributed energy resources at the distribution level, and the electrification of private transportation are considered as suitable governmental policies to tackle some of the problems of advanced societies, such as reducing CO_2 emissions or increasing energy efficiency [Gellings, 2009]. In recent years, these solution concepts started to pose new challenges to the existing power grids, whose hierarchical, centrally-controlled structure has remained unchanged for a century. For example, the exploitation of renewable sources such as solar or wind may be problematic due to their variable and intermittent nature, while the integration of distributed energy resources may cause congestion and atypical power flows that threaten system's reliability [Mohd et al., 2008].

In this context, energy consumption prediction for different time horizons (e.g., 1 hour ahead, 1 day ahead, 1 month ahead) and space scales (e.g., distribution transformer, individual house-level meter) is becoming crucial for many applications, such as frequency and voltage regulation, demand response (to estimate customer's baseline [Wijaya et al., 2014d]), and autonomous emergency management [Moslehi and Kumar, 2010]. While long-term load forecasting (1-10 years ahead) is important for planning both, transmission and distribution networks, short-term load forecasting (hours to days ahead) is important for the demand response, online scheduling, and security functions of an energy management system. In this chapter, we use the terms *energy consumption* (or *demand*) and *load* interchangeably.

Overview of Contributions Since energy consumption behavior might vary among households, a feature that are relevant for one house might not be relevant for others. Additionally, we have a large number of houses. Thus, feature selection has to be done automatically. To this end, we first build a (large) feature universe, and then automatically determine the relevant features for each house using the Correlation-based Feature Selection [Hall, 1999], which selects subset of features set that are highly correlated with the response variable while having low inter-correlation between each other (see Section 4.3.1).

As we have described in Section 2.1.2, many techniques for energy consumption prediction have been inspired by research on statistical and machine learning, from Linear Regression [Hong, 2010, Papalexopoulos and Hesterberg, 1990], ARMA [Huang and Shih, 2003, Taylor, 2010], and Generalized Additive Models [Ba et al., 2012, Fan and Hyndman, 2012] to Neural Networks [An et al., 2013, Hippert et al., 2001, Khotanzad et al., 1997] and Support Vector Regression [Chen et al., 2004, Sapankevych and Sankar, 2009]. However, these techniques have been typically used at very large space scales, such as predicting the electrical load of a market segment serving thousands of customers or even an entire country. In this chapter, we show that these algorithms can also be used to forecast households' consumption and improve the benchmark by around 20%–24% (see Section 4.3.2 and Section 4.4). Looking at prediction results, however, load forecasting at the household level remains a hard problem.

Misiti et al. [2010] studied the effect of forecasting clusters of industrial customers to predict their aggregate demand using wavelet-based clustering.[1] Alzate and Sinn [2013] used kernel spectral clustering and consider a mix of residential customers and small/medium enterprises. Interestingly, although Misiti et al., Alzate and Sinn, and our work focus on different customer bases and use different forecasting and clustering algorithms, all conclude that clustering customers and then forecasting each cluster separately could indeed improve aggregate forecasts. Additionally, our clustering objective is clear, targeting a specific property of the resulting cluster (see Section 4.5.1), and we continue by investigating how the improvement provided by this Cluster-based Aggregate Forecasting strategy depends not only on the number of clusters, but also on the size of the customer base. That is, the larger the customer base, the higher the improvement (see Section 4.5.2). Thus, our finding offers additional insight to the practitioners who wish to implement this strategy in the real world.

4.2 Dataset and Evaluation Metrics

4.2.1 Dataset

We use the Irish CER dataset, which contains energy consumption measurements of around 5,000 customers over 1.5 years [The Commission for Energy Regulation (CER) and the ISSDA, 2012] (see also Section 3.5.1). The measurements started in July 2009 and ended in December 2010, and recorded energy consumption in kWh every 30 minutes. We choose residential customers

[1] They refer to this approach as *disaggregated load forecasting*. To avoid confusion with the individual load forecasting, we use the term Cluster-based Aggregate Forecasting instead.

that belong to the control group and have no missing values. This results in the selection of 782 customers. The measurements are aggregated into hourly timeslots. For all results presented in this chapter, we use the first year (from July 2009 to June 2010) as the training set, and the remaining 6 months (from July 2010 to December 2010) as the test set.

4.2.2 Evaluation Metrics

In the literature, there are three widely used metrics to evaluate the accuracy of a forecasting algorithm: the Mean Absolute Percentage Error (MAPE), the Mean Absolute Error (MAE), and the Root Mean Square Error (RMSE). Given a time series $S = \{s_1, s_2, \ldots, s_n\}$ of observed consumption values and the estimation produced by forecasting algorithm $\widehat{S} = \{\widehat{s}_1, \widehat{s}_2, \ldots, \widehat{s}_n\}$, the MAPE is defined as:

$$\mathrm{MAPE}(S, \widehat{S}) = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{s_t - \widehat{s}_t}{s_t} \right| \tag{4.1}$$

The MAPE is a quite intuitive metric. However, it has a major drawback, i.e., it is not robust to the division by values approaching zero. Many households in the dataset have zero consumption on certain time slots, which makes the MAPE undefined, Furthermore, it is quite common to have households with very small consumption values, which makes the MAPE very large, approaching infinity.

Unlike the MAPE, the MAE and the RMSE do not suffer from the division by values approaching zero, since the MAE is defined as

$$\mathrm{MAE}(S, \widehat{S}) = \frac{1}{n} \sum_{t=1}^{n} (s_t - \widehat{s}_t), \tag{4.2}$$

and the RMSE is defined as

$$\mathrm{RMSE}(S, \widehat{S}) = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (s_t - \widehat{s}_t)^2}. \tag{4.3}$$

However, they are scale-dependent metrics. Since the average hourly consumption of households in the dataset varies between 0.05 kWh and 3.83 kWh, we need scale-independent metrics to aggregate the forecasting error of these different households. Moreover, scale-independent metrics can be useful to compare not only the forecasting error of different households, but also the forecasting error of different temporal aggregations or customer groups.

To this end, we suggest to use other metrics that are both, scale-independent, and robust to the division by values approaching zero, namely the Normalized Mean Absolute Error (NMAE) and the Normalized Root Mean Square Error (NRMSE). The NMAE is defined as

$$\mathrm{NMAE}(S, \widehat{S}) = \frac{\mathrm{MAE}(S, \widehat{S})}{||S||_1} = \left(\sum_{t=1}^{n} |s_t - \widehat{s}_t| \right) \Big/ \left(\sum_{t=1}^{n} |s_t| \right), \tag{4.4}$$

and the NRMSE is defined as

$$\mathrm{NRMSE}(S, \widehat{S}) = \frac{\mathrm{RMSE}(S, \widehat{S})}{||S||_2} = \sqrt{\left(\sum_{t=1}^{n} (s_t - \widehat{s}_t)^2 \right) \Big/ \left(\sum_{t=1}^{n} s_t^2 \right)}. \tag{4.5}$$

While one zero measurement is enough to make the MAPE undefined (or approaches infinity), all measurements need to be zero to make the NMAE or the NRMSE undefined.

4.3 Forecasting Models

4.3.1 Features

There are two important challenges in selecting features for residential electricity load forecasting. First, different houses might have different energy consumption behavior. Thus, features that are relevant to one house might not be relevant to other houses. Second, we have a large number of houses. Therefore, feature selection should be done automatically.

To solve both challenges, we first build a (large) feature universe and then apply a feature selection algorithm to select features that are relevant to each house. We consider both, historical load and contextual features. To forecast the load at time (or hour) t, for 1 hour ahead forecasting, we consider the historical load data from time $t-1$ to $t-336$, i.e., $\{s_{t-1}, s_{t_2}, \ldots, s_{t-336}\}$.[2] While for 24 hour ahead forecasting, we consider the historical load data from time $t-24$ to $t-336$ (since the historical load data from time $t-1$ to $t-23$ is not available in this case).

The CER dataset does not contain any information about the house or the persons who live in the house. Thus, for contextual features, we consider day of week, hour of day, and weather information. Since there is no information about the city/location of each house, we crawl the historical weather data of the three biggest cities in Ireland, i.e., Dublin, Cork, and Limerick.[3] We use 48 hours historical temperature and humidity data,[4] from time $t-1$ to $t-48$ for 1 hour ahead forecasting, and from time $t-24$ to $t-71$ for 24 hour ahead forecasting. Additionally, we also use include the mean and the median of those three cities to the feature set.

Up to this point, our feature universe contains approximately 800 variables. Next, we apply Correlation-based Feature Selection (CFS) to each house. This method selects subset of features that are highly correlated with the response variable while having low inter-correlation between each other [Hall, 1999]. As a result, we obtain a (much) reduced subset of relevant features for each house.

4.3.2 Learning Algorithms

Various learning algorithms have been used to forecast large-scale electricity demand. Recent literature suggests Support Vector Regression (SVR) as one of the most effective models to forecast future energy consumption [Chen et al., 2004, Sapankevych and Sankar, 2009]. Other well established methods are Linear Regression and Multi-Layer Perceptron (MLP). In this section, we briefly describe our model setup.

Linear Regression configuration

A linear model to predict the load at time t is defined as:

$$y = \boldsymbol{\theta}^T \boldsymbol{x} + \epsilon \tag{4.6}$$

where $\boldsymbol{\theta}$ is the vector of coefficients, $\boldsymbol{x}$ is the feature vector, and ϵ is the error term. We estimate the coefficients and the error term of the linear model using ridge regression (other methods, of course, can also be used).

[2] Of course, longer time duration can also be considered here, in the price of memory and computation cost.

[3] We obtained the weather-related data from http://www.wunderground.com.

[4] Apart from temperature, humidity has also been used in real-world implementation to forecast electricity demand. See, e.g., [KEMA, Inc., 2011].

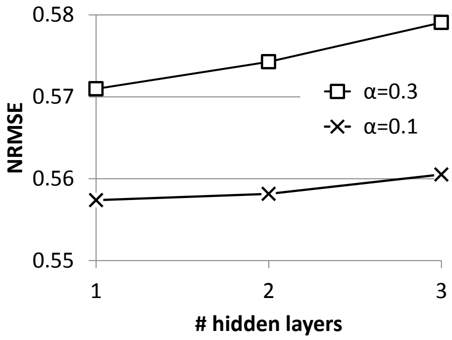


Figure 4.1: MLP model evaluation (using NRMSE) using different number of hidden layers and learning rates α on randomly chosen 25 households. The lower the better. In the end, we use one hidden layer and $\alpha = 0.1$.

C \ γ	0	0.01	0.1	1
1.E+00	0.69	0.62	0.59	0.57
1.E+01	0.62	0.59	0.58	0.57
1.E+02	0.59	**0.58**	0.57	0.58
1.E+03	0.58	0.58	0.57	0.63
1.E+04	0.58	0.57	0.58	0.82
1.E+05	0.57	0.57		

(a)

C \ γ	0	0.01	0.1	1
1.E+00	0.02	0.03	0.02	0.02
1.E+01	0.03	0.02	0.02	0.20
1.E+02	0.02	**0.02**	0.11	0.02
1.E+03	0.02	0.02	0.02	0.02
1.E+04	0.02	0.02	0.02	0.04
1.E+05	0.02	0.11		

(b)

C \ γ	0	0.01	0.1	1
1.E+00	2.1	2.7	3.2	4.6
1.E+01	2.3	2.6	3.1	3.2
1.E+02	2.3	**2.4**	4.0	8.1
1.E+03	2.7	3.0	8.9	45
1.E+04	4.0	8.2	50	222
1.E+05	7.2	52		

(c)

Figure 4.2: SVR model evaluation for individual forecasting on the randomly chosen 25 households: (a) average NRMSE on the validation set given different C and γ, (b) standard deviation on the average, (c) average running time. The lower the better. In the end, we choose $C = 100$ and $\gamma = 0.01$. While there are some other settings which yield better NRMSE, they typically require considerably longer running time.

MLP configuration

We use one hidden layer with sigmoid activation functions. The output can be written as $y = \boldsymbol{W}_2 \times \Theta(\boldsymbol{W}_1 \cdot \boldsymbol{x} + \boldsymbol{B}_1) + \boldsymbol{B}_2$ where $\boldsymbol{x}$ is the input vector, y is the output value, $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, $\boldsymbol{B}_1$, and $\boldsymbol{B}_2$ are the coefficient matrices, and Θ is the sigmoid operator. Each component x_j of the input vector $\boldsymbol{x}$ is standardized, i.e., $x_j^* = (x_j - \mu_j)/\sigma_j$, where μ_j is the mean and σ_j is the standard deviation of the values in the jth dimension. To avoid overfitting, a validation set is constructed by randomly selecting 30% of the instances in the training set. The coefficient matrices are learnt using gradient descent, with learning rate of $\alpha = 0.1$ (see the evaluation of different hidden layers and learning rates in Figure 4.1). The stopping criterion is triggered when the error on the validation set (calculated after each epoch) has increased 20 times in a row.

SVR configuration

SVR is a regression method based on Support Vector Machine (SVM) that has been developed in 1996 by Vapnik (see also the tutorial by Smola and Schlkopf [2004]). In this work, we use the SVR implementation provided by the LIBSVM library developed by Chang and Lin [2011].

SVR must be provided with the SVM error cost C and a kernel function. For the kernel function, we use the RBF kernel, similar to [Chen et al., 2004]. Next, to find suitable values for

C \ γ	0.001	0.01	0.1	1	10
1.E-01	0.418	0.415	0.395	0.356	0.318
1.E+00	0.415	0.392	0.226	0.160	0.201
1.E+01	0.391	0.200	0.074	0.070	0.082
1.E+02	0.197	0.073	0.064	0.052	0.082
1.E+03	0.073	0.066	0.059	**0.045**	0.065
1.E+04	0.066	0.065	0.054	**0.045**	0.072
1.E+05	0.066	0.061	0.050	0.050	
1.E+06	0.065	0.057	0.047	0.063	
1.E+07	0.061	0.054	0.046	0.087	
1.E+08	0.063	0.110	0.188	0.149	

Figure 4.3: SVR model evaluation (measured by average NRMSE) for aggregate forecasting. The lower the better.

C and γ, we split the training set into two parts: a sub-training set and a validation set. The SVR is trained on the sub-training set, and evaluated on the validation set. For C we test a set of values $\{1, 10, 10^2, 10^3, 10^4, 10^5\}$, while for γ we test a set of values $\{0, 0.01, 0.1, 1\}$.

For individual load forecasting, we find that different values of C and γ do not result in significant NRMSE differences (see Figure 4.2a and 4.2b) . However, they strongly affect the computation time, which dramatically increase when $C \geq 1000$ or $\gamma \geq 0.1$ (see Figure 4.2c). Thus, for individual forecast, we use $C = 100$ and $\gamma = 0.01$. On the other hand, for aggregate forecast, different settings of C and γ result in significant differences in terms of NRMSE (see Figure 4.3). We found that $C = 1000$ and $\gamma = 1$ is the best setting.

4.4 Individual Forecasting

In addition to features and learning algorithms, we also explore pth root transformation. That is, instead of modeling the response variable (s_t) as is, we model its pth root ($s_t^{1/p}$), and then transform the forecasted value back to its original dimension by raising it to the pth power ($(\widehat{s}_t)^p$). Since the distribution of household energy consumption are skewed to the left toward zero, pth root transformation could help to make it more normal and easier to model.

Tables 4.1 and 4.2 show the performance of Linear Regression (`LR`), Multi-Layer Perceptron (`MLP`), and Support Vector Regression (`SVR`) using the setting described in Section 4.3. Both tables show that the pth root transformation mostly improves the NMAE of the models. Additionally, as a comparison to the three models above, we use *persistence forecast* as the benchmarks, i.e., the load of the previous hour (s_{t-1}) for the 1 hour ahead forecasting, and the load for the same hour of the previous day (s_{t-24}) for the 24 hour ahead forecasting. Although household-level forecasting is a difficult problem, we show that it is possible to improve the prediction by around 20% – 24% compared to the benchmark. Moreover, the improvements provided by the three learning algorithms for the 24 hour ahead forecasting are consistently higher than that of 1 hour ahead, which shows the greater advantage of using the learning algorithms (rather than the benchmarks) on a longer forecasting horizon.

Additionally, as a comparison to the three models above, we employ time series analysis method, namely Seasonal ARIMA (or `SARIMA`). Before using `SARIMA`, however, one need to properly identify the order of the autoregressive, integrated, and moving average terms (for both, the seasonal and non-seasonal parts). Similar to the challenges that we face in the feature selection procedure, there are two important challenges. First, since different households might have different energy consumption behavior, we need to identify the right orders for each

Table 4.1: Average NRMSE and NMAE (with its 95% confidence interval) of LR, MLP, SVR for 1 hour ahead load forecasting at the level of the individual customer. Benchmark (bm) is s_{t-1}. The numbers in parentheses show the improvements compared to the benchmark. Root transformation (${s_t}^{1/p}$, with $p > 1$) can be used to improve NMAE.

		$p = 1$	$p = 2$	$p = 4$
NRMSE	bm	0.694 ± 0.010	0.694 ± 0.010	0.694 ± 0.010
	LR	**0.557 ± 0.007 (19.7%)**	0.562 ± 0.007 (19.1%)	0.571 ± 0.007 (17.7%)
	MLP	0.575 ± 0.008 (17.2%)	0.569 ± 0.007 (18.1%)	0.578 ± 0.008 (16.6%)
	SVR	0.573 ± 0.008 (17.4%)	0.571 ± 0.007 (17.8%)	0.572 ± 0.007 (17.6%)
NMAE	bm	0.562 ± 0.010	0.562 ± 0.011	0.562 ± 0.011
	LR	0.495 ± 0.009 (11.9%)	0.461 ± 0.007 (17.9%)	0.456 ± 0.007 (18.9%)
	MLP	0.535 ± 0.014 (4.7%)	0.477 ± 0.009 (15.0%)	0.468 ± 0.008 (16.7%)
	SVR	0.461 ± 0.007 (17.9%)	**0.448 ± 0.007 (20.2%)**	0.452 ± 0.007 (19.5%)

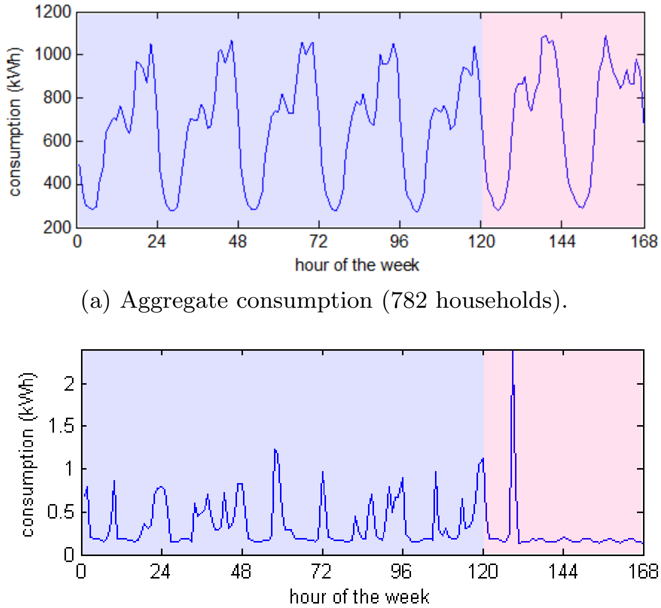
Table 4.2: Average NRMSE and NMAE (with its 95% confidence interval) of LR, MLP, SVR for 24 hour ahead load forecasting at the level of the individual customer. Benchmark (bm) is s_{t-24}. The numbers in parentheses show the improvements compared to the benchmark. Root transformation (${s_t}^{1/p}$, with $p > 1$) can be used to improve NMAE.

		$p = 1$	$p = 2$	$p = 4$
NRMSE	bm	0.802 ± 0.010	0.802 ± 0.010	0.802 ± 0.010
	LR	**0.607 ± 0.007 (24.3%)**	0.613 ± 0.007 (23.6%)	0.623 ± 0.008 (22.3%)
	MLP	0.633 ± 0.008 (21.2%)	0.630 ± 0.008 (21.5%)	0.638 ± 0.008 (20.5%)
	SVR	0.628 ± 0.008 (21.8%)	0.628 ± 0.008 (21.8%)	0.629 ± 0.008 (21.7%)
NMAE	bm	0.661 ± 0.011	0.661 ± 0.011	0.661 ± 0.011
	LR	0.555 ± 0.010 (16.0%)	0.515 ± 0.008 (22.1%)	0.507 ± 0.008 (23.3%)
	MLP	0.601 ± 0.016 (9.1%)	0.541 ± 0.010 (18.2%)	0.527 ± 0.009 (20.2%)
	SVR	0.512 ± 0.008 (22.6%)	**0.501 ± 0.008 (24.2%)**	0.508 ± 0.008 (23.1%)

Table 4.3: Average NRMSE and NMAE (with its 95% confidence interval) of Seasonal ARIMA (SARIMA) for 1 hour and 24 hour ahead individual load forecasting. Benchmark (bm) is s_{t-1} for 1 hour ahead forecasting and s_{t-24} for 24 hour ahead forecasting. We use the pth root transformation (${s_t}^{1/p}$), with $p = 2$. The numbers in the parentheses show the improvements compared to the benchmarks.

		1 hour ahead, $p = 2$	24 hour ahead, $p = 2$
NRMSE	bm	0.694 ± 0.010	0.802 ± 0.010
	SARIMA	0.582 ± 0.007 (16.1%)	0.674 ± 0.007 (19.1%)
NMAE	bm	0.562 ± 0.010	0.661 ± 0.011
	SARIMA	0.485 ± 0.008 (13.7%)	0.598 ± 0.009 (17.9%)

household (i.e., the orders that are suitable for one household might not be suitable for others). Second, we have a large number of households. Thus, the identification procedure need to be done automatically. To this end, we apply the stepwise model space exploration algorithm outlined in [Hyndman and Khandakar, 2008] to each household. Starting from a small set of models, the algorithm iteratively explore the "neighbors" of the best model found so far. The algorithm stops when it cannot find a model better than the current best model. We show the results in Table 4.3. Although SARIMA provides significantly better forecasts than benchmarks, it does not outperform LR, MLP, or SVR in this case (cf. Table 4.1 for 1 hour ahead, and Table 4.2 for 24 hour ahead, with $p = 2$).



(a) Aggregate consumption (782 households).

(b) Example consumption of a household (id 1002).

Figure 4.4: A sample of hourly energy consumption from the CER dataset, from Monday, 2009-09-07 to Sunday, 2009-09-13.

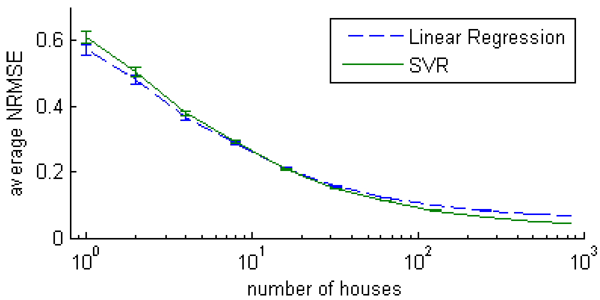


Figure 4.5: The NRMSE of `LR` and `SVR` for 1 hour ahead forecasting (the lower the better). Forecasting error decreases as the aggregation size increases.

4.5 Aggregate Forecasting

In order to provide an aggregate forecast of a set of individually-monitored households, it is possible to define two extreme strategies: (1) aggregate the energy consumption of all households into one time series (the aggregate consumption), then forecast the aggregate consumption, and (2) forecast the energy consumption of each household separately, then aggregate the forecasts. Since the patterns in aggregate consumption are more regular than that of individual consumption (see also Figure 4.4), intuitively, strategy (1) should outperform strategy (2). Figure 4.5 shows clearly that the forecasting error decreases as the aggregation size increases.

In this section, we evaluate an alternative strategy (3), where we segment the households into k clusters, aggregate the energy consumption of the households in each cluster, forecast each cluster separately, and finally aggregate the k forecasts into one aggregate forecast. Strategy (1) and (2) can also be seen as some special cases of strategy (3), where $k = 1$ and $k = N =$ total customers, respectively. We refer to strategy (3) as the Cluster-based Aggregate Forecasting (CBAF). The contributions of this section are: (i) we provide clustering algorithms to form clusters with some predefined/targeted characteristics (see Section 4.5.1), whereas previous works offer only little interpretation to the characteristics of the resulting clusters, (ii) we find that the improvement provided by the CBAF strategy depends not only on the number of clusters, but also on the size of the customer base (see Section 4.5.2).

4.5.1 Clustering Algorithms

In order to investigate the effectiveness of CBAF, we define several clustering methods with clear objective, targeting a specific property of the resulting clusters:

- `Max-AC`: This method aims to maximize the absolute auto-correlation of the energy consumption of the clusters. More specifically, this method uses the greedy clustering technique proposed in Algorithm 4.1 to find clusters such that the absolute auto-correlation of the load of each cluster is maximized. Let $ac(S)$ be the average auto-correlation (up to a certain lag) of time series S.[5] In addition, we define a *cluster* as a set of customers, and S_c as the aggregate consumption time series of cluster c. Then this method uses Algorithm 4.1 by defining
$$\Phi(c, x) = |ac(S_{c\cup\{x\}}) - ac(S_c)|,$$
where x is a customer. As a consequence, customer x is assigned to a cluster where x provides the highest improvement to the absolute auto-correlation of the clusters' energy consumption.

- `Min-Stdev`: This method aims to minimize the fluctuation in the clusters' energy consumption, which often becomes the main challenge to predict. In particular, it aims to minimize the standard deviation of the clusters' energy consumption. Let $sd(S)$ be the standard deviation of time series S. As in the `Max-AC` case, we define a cluster as a set of customers, and S_c as the aggregate consumption time series of cluster c. Then this method uses Algorithm 4.1 by defining
$$\Phi(c, x) = (sd(S_c) - sd(S_{c\cup\{x\}})) \cdot |c|,$$
where x is a customer. As a consequence, customer x is assigned to a cluster where x minimizes the standard deviation of the cluster's aggregate consumption. Note that in the evaluation function Φ, we multiply the standard deviation difference by $|c|$ so as to have a weighted difference, with respect to the size of the original cluster c (before x is added).

- `Max-Sim`: This method aims to maximize the similarity among customers within a cluster. Unlike previous two methods, here we apply kMeans clustering algorithm to customer's 24-hour load profiles, where each hour is characterized by the distribution (or histogram) of the amount of energy consumed in that hour. More specifically, for each hour, we define a feature vector of length 21. For the first 20 elements, the ith element is the frequency of consumption between $(i-1)\times 0.5$ kWh and $i\times 0.5$ kWh. The 21[st] element is the frequency of consumption greater than 10 kWh. Finally, we apply kMeans on the customers' feature vectors, where each feature vector of a customer is of length $24 \times 21 = 504$.

[5] In our implementation, we compute the auto-correlation up to lag 168 (or, 1 week preceding the target time). Other lags, however, can also be used.

Algorithm 4.1: Generic greedy clustering algorithm

Input: number of clusters k, customer set X
Output: cluster configuration $C = \{c_1, \ldots, c_k\}$

```
{x_1, ..., x_k} ← draw randomly k customers from X
for i ∈ {1, ..., k} do  c_i ← x_i                    /*initialization*/
while X ≠ ∅ do
    x ← draw randomly a customer from X
    c* ← arg max_{c∈C} Φ(c, x)
    c* ← c* ∪ {x}
return C
```

- `Random`: Each customer is randomly assigned to any of the clusters with the equal probability.

In this experiment, we enlarge our dataset to include all residential customers who have no missing values. This results in the selection of 3,639 customers. Figures 4.6 shows the NRMSE, the NMAE and the MAPE of `LR`, `MLP` and `SVR`, for a different number of clusters k. When $k = 1$, all customers are aggregated into a single cluster and a single prediction is performed. As k increases, more clusters are created (k clusters to be precise), and the consumption of each cluster is forecasted separately. The forecasts are then aggregated into a single aggregate forecast. Note that, $k = 1$ represents strategy (1), $k = N = 3639$ strategy (2), and $1 < k < N$ the CBAF strategy, which is the focus of this section. We do not show the forecasting result beyond $k = 128$ as the error continues to increase beyond that of $k = 1$. This fact shows clearly that strategy (1) outperforms strategy (2). Additionally, there are some $1 < k < N$ for which the forecasting error is lower than that of $k = 1$. This fact confirms that CBAF indeed can be used to improve the accuracy of aggregate forecasting.

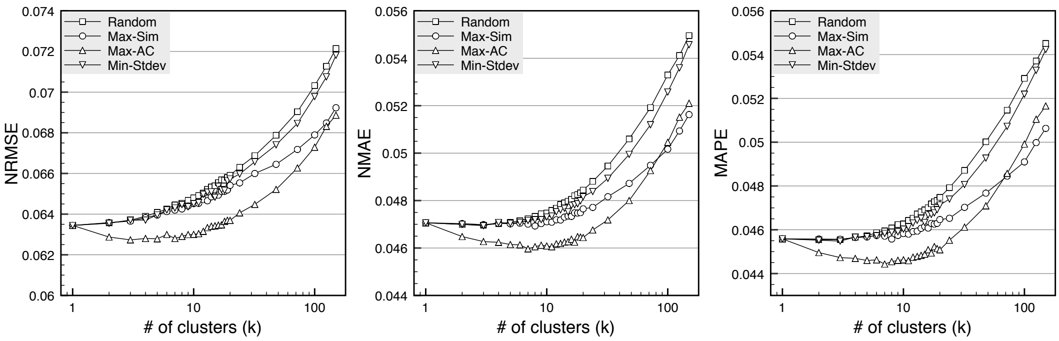
Interestingly, all clustering methods that we introduced (including `Random`) seem to be able to provide a lower forecasting error than that of strategy (1). Although in some cases `Max-AC` provides the lowest error curve as it aims to maximize the auto-correlation of the energy consumption within the clusters,[6] the accuracies of these clustering methods are often marginally different. Therefore, choosing one clustering method against the others (or implementing the CBAF strategy) in a real-world scenario needs a more careful analysis, in the sense that we need to consider whether the advantage brought by a particular clustering method is greater than the cost of implementing it.

4.5.2 The Impact of the Size of the Customer Base

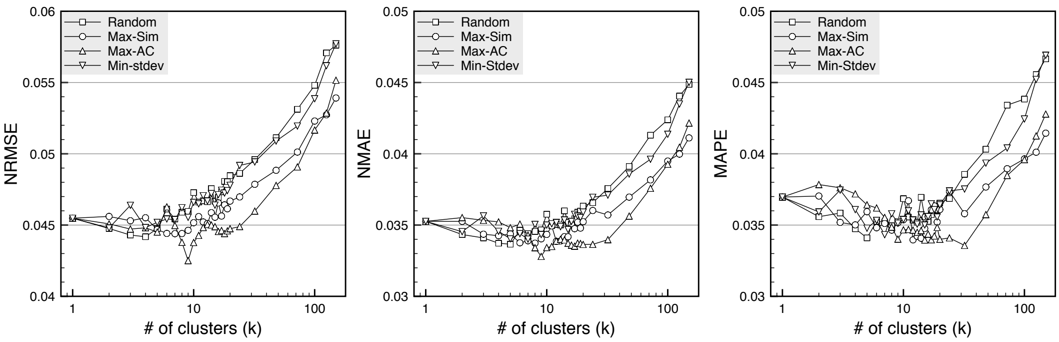
Hitherto one might think that the improvement obtained by CBAF depends on the number of clusters, k. While this insight has been confirmed by Figure 4.6, there is more to it than that since it turned out that the size of the customer base also plays an important role in the improvement. We repeat the experiments using different sized customer bases: 500, 1,000, and 2,000 (drawn randomly from the original dataset of 3,639 customers). Note that, for the same k, a different size of customer bases implies different cluster sizes.

Figure 4.7 shows the improvement gained by `SVR` when we perform CBAF on different size of customer bases. While there is almost no positive improvement in the case of 500 customers (no

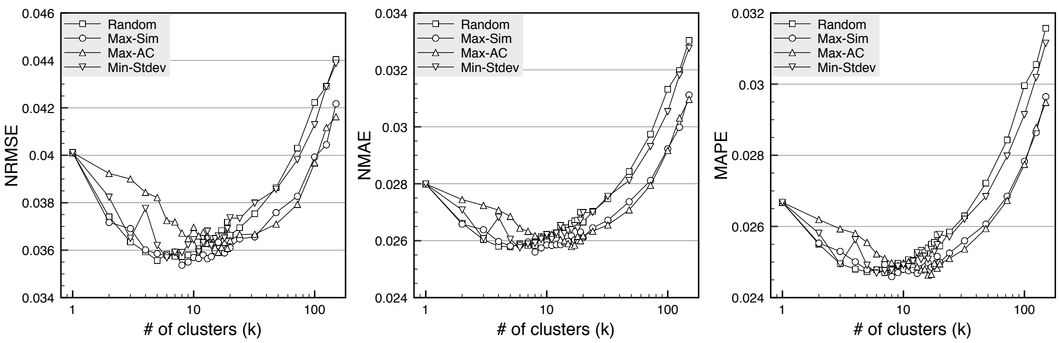
[6] Time series with higher auto-correlation is typically easier to forecast since it shows greater relationship between the current and the past values.



(a) LR



(b) MLP



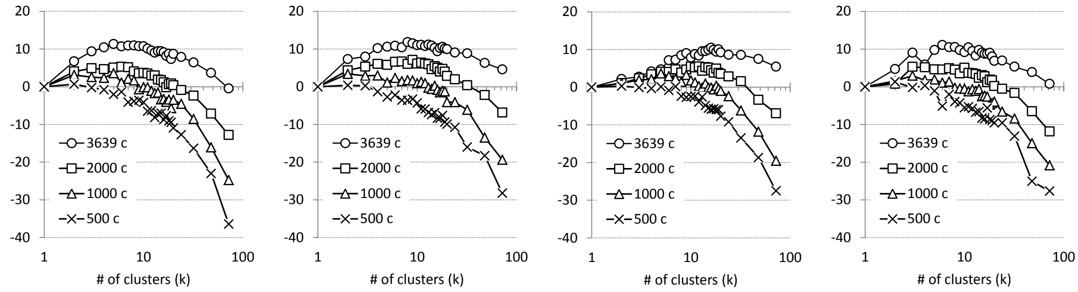
(c) SVR

Figure 4.6: The NRMSE, the NMAE and the MAPE for a different number of clusters k (the lower the better). Total number of customers, $N = 3639$. The best accuracy is obtained when $1 < k < 3639$, which shows the effectiveness of CBAF.

matter which clustering method is used), some may be noticed in the case of 1,000 customers or more. In general, the improvement increases with the size of the customer base.

If we assume that a "good" forecast models the true observation and a white noise (zero mean and finite variance), then it means that combining several good forecasts from several clusters into one aggregate forecast could neutralize the white noise. Thus, there is a trade-off between the size and the number of clusters. The size of the clusters should be big enough for the algorithm to deliver a reasonably good prediction,[7] but not too big that there are not enough

[7] See also Figure 4.5 about the relation between forecast accuracy and customer aggregation size.



(a) `Random` (b) `Max-Sim` (c) `Max-AC` (d) `Min-Stdev`

Figure 4.7: Percentage improvement in the NRMSE of the CBAF strategy (compared to the traditional aggregate forecast, $k = 1$) of 500, 1,000, 2,000, and 3,639 customers over a different number of clusters and clustering methods (the higher the better). The larger the customer set, the higher the improvement gained by CBAF.

clusters (hence, predictions) to cancel out the noise.

In addition, since the number of clusters, k, strongly influence the cluster size, one might wonder whether it is possible to set a priori the best value for k. Because characteristics of a customer base vary from one to another, the right k should be determined using *cross validation* on the training set. Apart from that, the experiment suggests that the improvement offered by CBAF will increase further as we incorporate more and more customers (due to the possibility to increase both, the number and the size of the clusters).

4.6 Summary and Discussion

In this chapter we evaluate different machine learning algorithms (`LR`, `MLP`, and `SVR`) for short-term individual and aggregate forecasting (1 hour and 24 hour ahead) of residential electricity consumption. Additionally, to measure the forecasting accuracy at the household level effectively, we use evaluation metrics that are scale-independent and robust to values approaching zero, namely the NRMSE and the NMAE.

Individual forecasting, in general, is a challenging task (with NRMSE around 0.6 and NMAE around 0.4–0.5). Aggregate forecasting, however, have better accuracy (with NRMSE around 0.04 and NMAE around 0.03). While in individual forecasting we improve the benchmark by around 20%, using similar techniques, in aggregate forecasting we improve the benchmark by around 74%–78% (NRMSE and NMAE of the benchmark in aggregate forecasting are 0.156 and 0.130 respectively).

Although `MLP` and `SVR` are more sophisticated than `LR`, in individual forecasting, their forecasting performances are not significantly better than `LR` (especially with pth root transformation, where $p = 2$ or $p = 4$). In aggregate forecasting, however, `SVR` is significantly better than `LR` (see, e.g, Figure 4.6 where $k = 1$). Therefore, in a real-world scenario, one should consider the trade-off between the advantage brought by a more sophisticated model and the cost to implement and maintain it.

In addition, we proposed a generic algorithm to segment customers according to a predefined/targeted objective. We showed its usefulness by forming clusters that (1) maximize the auto-correlation and (2) minimize the standard deviation of the clusters' energy consumption. Additionally, we found that the improvement provided by the CBAF strategy depends not only on the number of clusters, but also on the size of the customer base. More specifically, CBAF improves traditional aggregate forecasting when the size of the customer base is above a certain threshold. Conversely, no improvement is achieved when the size of the customer base is below this threshold, no matter which clustering methods is applied. In general, however, the larger the size of the customer base, the higher the improvement offered by CBAF.

5 Modeling Uncertainty of Future Demand

Generalized Additive Models (GAM) are an increasingly popular class of regression models to forecast electricity demand, due to their high accuracy, flexibility and interpretability. However, the residuals of the fitted GAM are typically heteroscedastic and leptokurtic caused by the nature of energy data. This chapter proposes a novel approach to estimate the time-varying conditional variance of the GAM residuals, which we call the GAM^2 algorithm. It allows utility companies and network operators to assess the uncertainty of future electricity demand and incorporate it into their planning processes. The basic idea of our algorithm is to apply another GAM to the squared residuals to explain the dependence of uncertainty on exogenous variables. Empirical evidence shows that the residuals rescaled by the estimated conditional variance are approximately normal. We combine our modeling approach with online learning algorithms that adjust for dynamic changes in the distributions of demand. We illustrate our method by a case study on data from Réseau de transport d'électricité, the operator of the French transmission grid.

The work for this chapter was carried out during the author's internship at IBM Research, Ireland. This chapter has also been presented in the AAAI-15 Workshop on Computational Sustainability [see Wijaya et al., 2015].

5.1 Introduction

Forecasting electricity demand is a key instrument in operational and planning processes of electric utilities. In this chapter, we focus on short-term forecasts (with a forecast horizon of 24-48 hour ahead), which are required, e.g., by electricity suppliers to bid generation/load into electricity markets, and by network operators for day-ahead outage planning. As we have outlined in Section 1.1, however, due to an ever growing population and carbon emission, the electricity sector has been changing dramatically in the last few years, which made electricity demand forecasting even more challenging. Electric vehicles (EVs) also pose a new challenge to electricity grids by drawing a large amount of energy in a very short time.[1] Moreover, the

[1] Charging one EV can consume 32 kWh, which is comparable to one household's daily consumption [Hess et al., 2012, Ramchurn et al., 2012].

integration of renewable energy sources to the grid, as an effort to reduce our dependency on fossil fuels, adds additional complexity to efforts at balancing supply and demand, since their generation is intermittent and unpredictable.

Therefore, it is imperative for utility companies and networks operators to forecast not only the conditional expectation (as in Chapter 4) but also the uncertainty of the future demand, compute the risk/benefit associated with it, and incorporate it into their planning process. However, a large body of literature so far has been focused on single-valued, "point" forecast to estimate the conditional mean of the future demand. Compared to the point forecast, forecasting uncertainty is more challenging since it requires us to estimate the entire distribution of the future demand. In this chapter, we propose a novel approach to address the problem by modeling the time varying conditional mean and variance of the future demand using Generalized Additive Models (GAM).

Probabilistic Forecasting Gneiting and Katzfuss [2014] provide a good overview on probabilistic forecasting. It has been an increasingly important direction to model uncertainty in various fields, such as healthcare [Jones and Spiegelhalter, 2012], politics [Montgomery et al., 2012], weather [Palmer et al., 2005, Warner, 2011], and finance [Groen et al., 2013]. Probabilistic forecasting has also drawn more and more interest in the energy domain, e.g., for long-term [Hong et al., 2014b, Hyndman and Fan, 2010] and short-term electricity demand forecasting [Fan and Hyndman, 2012], solar power [Bacher et al., 2009], and wind power forecasting [Wytock and Kolter, 2013]. While this chapter was being written, even a public competition has been dedicated to probabilistic energy forecasts.[2] Probabilistic forecasts typically rely on computationally expensive approaches such as simulation, bootstrapping, or ensemble forecasting, which all require running multiple forecast to generate prediction intervals. In contrast, our approach constructs prediction intervals directly using the estimated conditional mean and variance. Thus, it does not require multiple runs of forecasting simulations/scenarios.

Generalized Additive Models Among other forecasting methods in the literature (see Section 2.1.2), GAM has been increasingly popular due to its accuracy, flexibility, and interpretability [Hastie et al., 2009, Wood, 2006]. Those advantages have made GAM attractive also for energy analytics where understandability is a key criteria for a model to be deployed by utility companies. While other (accurate) models are typically opaque and difficult to interpret, GAM consists of transfer functions that are easy to understand.[3] Additionally, the "transparency" of GAM opens possibilities for practitioners to discover new insights about the relationship between some exogenous and response variables, or the other way around, to spot a potential overfitting when a relationship does not conform a well-established physical law.

Overview of Contributions The contributions of this chapter are as follows. We propose a novel GAM^2 (or *GAM squared*) algorithm to forecast uncertainty in electricity demand (Section 5.2.1). First, we use a GAM to model the conditional expectation of the demand. The residuals of the fitted GAM, however, are typically heteroscedastic and leptokurtic due to the nature of energy data. To this end, we apply a second GAM to the squared residuals to explain the dependence of uncertainty on exogeneous variables and estimate the (time-varying) conditional variance. Under normality assumptions, the estimated conditional mean and variance allow us to construct prediction intervals. Although we showcase our method specifically for electricity demand, it can also be applied to other domain, as long as the rescaled residual is

[2] See GEFCom2014, `http://www.drhongtao.com/gefcom`.

[3] See, e.g., Figure 5.2.

(approximately) normal. Additionally, we propose a novel algorithm that adjust the prediction intervals to the non-stationary nature of electricity demand (Section 5.3). Finally, we illustrate the effectiveness of our approach on real electricity demand data provided by Réseau de transport d'électricité, France (Section 5.4).

5.2 Modeling Uncertainty

5.2.1 The GAM^2 Algorithm

We use the following general regression model for electricity demand:

$$Y_t = \mu(x_t) + \sigma(x_t)\epsilon_t \tag{5.1}$$

where y_t is the demand at time t, x_t is a vector of covariates (e.g., weekday, time of day, temperature, etc) and ϵ_t is the error, which we assume to have zero mean and unit variance. Note that under this assumption, the conditional mean and variance of Y_t (given x_t) are obtained by

$$\begin{aligned} \mathbb{E}[Y_t] &= \mu(x_t), \\ \mathrm{Var}(Y_t) &= \sigma^2(x_t). \end{aligned}$$

In practice, the functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are unknown and need to be estimated from empirical data. A wide range of methods has been explored in the literature for modeling the conditional mean demand function $\mu(\cdot)$, including Artifical Neural Networks, Support Vector Regression, seasonal time series models, and semi-parametric regression. In this chapter, we use GAM and express the conditional mean function as

$$F(\mu(x_t)) = \sum_{i=1}^{I} f_i(x_t),$$

where $F(\cdot)$ is a link function (e.g., the identity or logarithm function). The $f_i(\cdot)$ are called *transfer functions* and have the following form:

$$f_i(x_t) = \mathbf{1}_{(x_t \in A_i)} \beta_i^T b_i(x_t). \tag{5.2}$$

Here $\mathbf{1}_{(\star)}$ denotes the indicator function which returns 1 if the expression $\star$ is evaluated to true and 0, otherwise, and $b_i(\cdot)$ is a vector of *basis functions* (we mostly use cubic b-splines), typically depending on one or two continuous variables in the vector of covariates x_t. The indicator function allows for "switching on/off" transfer functions, e.g., to model the effect of the time of day depending on the weekday. See the experimental sections for examples.

We learn the GAM model using the `mgcv` package in `R`. This gives us an empirical fit $\widehat{\mu}(\cdot)$ of the conditional mean function $\mu(\cdot)$. In order to forecast the conditional mean at time t, given the vector of covariates x_t, we calculate

$$\widehat{y}_t = \widehat{\mu}(x_t). \tag{5.3}$$

The forecasting accuracy can be evaluated by considering the errors $\widehat{y}_t - y_t$ (or statistics thereof) on a held-out part of the training set.

Next, we focus on the uncertainty term $z_t := \sigma(x_t)\epsilon_t$ in Eq. 5.1. For now, our only assumption on the random variable ϵ_t is that it has zero mean and unit variance. We will discuss further assumptions (and their empirical justifications) below. For modeling the conditional variance function $\sigma^2(\cdot)$, we use another GAM model,

$$G(\sigma^2(x_t)) \quad = \quad \sum_{j=1}^{J} g_j(x_t).$$

where G is a link function and the transfer functions g_j have the same form as the f_i in Eq. 5.2. Note that the functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ may depend on different subsets of the covariates x_t. We will see examples in the experimental sections.

In order to fit a function $\widehat{\sigma}^2(\cdot)$ to empirical data, we proceed as follows:

1. We fit a function $\widehat{\mu}(\cdot)$ for the conditional mean demand, as explained above.
2. We calculate empirical residuals $\widehat{z}_t = \widehat{y}_t - y_t$.
3. We fit a GAM model to the squared empirical residuals $\widehat{z}_t^2$, again using the `mgcv` package in `R`.

Given the covariates x_t, we use this model to compute the estimate $\widehat{\sigma}^2(x_t)$ of the conditional variance. Note that a similar two-stage estimation procedure, based on non-parametric regression, was proposed in [Fan and Yao, 1998]. Our modeling approach takes into account the dependency of uncertainty on exogenous variable. In particular, the variance of electricity demand is typically higher during peak periods such as in the evening or on days with extreme temperatures. In statistical terms, such time-varying variance effects are known as heteroscedasticity. Normalizing $\widehat{z}_t$ by $\widehat{\sigma}(x_t)$ can be thought of a way to scale the empirical residuals such that the rescaled version has zero mean and one standard deviation.

Distributional assumptions So far we have been only assuming zero mean and unit variance of ϵ_t. If we assume normality, the conditional mean and variance uniquely parameterize the entire conditional distribution of y_t given x_t. As we show in the following paragraphs, this allows us to construct one- and two-sided prediction intervals. Without assuming normality, the conditional variance provides some information about the dispersion (and hence some quantification of uncertainty), which can be used, e.g., by applying Chebyshev's inequality to make probabilistic statements about deviations of the actual demand from its mean value. An empirical analysis in the experimental section shows statistical evidence that the normality assumption is indeed justified.

Another comment: in this chapter, we only focus on marginal distributions, i.e., characterizing the uncertainty at one particular time point t. In order to make statements about the uncertainty of demand in a time interval $T = \{t_1, ..., t_{|T|}\}$, one would have to make assumptions about the dependency structure of the process $\{\epsilon_t\}$, however, this goes beyond the scope of this work.

Prediction intervals Let $p \in [0, 1]$ and $q(p)$ be the quantile function of the standard normal distribution, i.e., if X has a standard normal distribution, then $Pr(X \leq q(p)) = p$. We construct the *one-sided prediction interval* at the $p \cdot 100$ *percentile* for Y_t as follows:

$$\phi_t^1(x_t, p) = \widehat{\mu}(x_t) + q(p) \cdot \widehat{\sigma}(x_t). \tag{5.4}$$

If the estimates $\widehat{\mu}(\cdot)$ and $\widehat{\sigma}^2(\cdot)$ are statistically consistent and ϵ_t follows a standard normal distribution, then we have $Pr(Y_t \leq \phi_t^1(x_t, p)) = p$. In the following, whenever the context is clear, we omit x_t and write $\phi_t^1(p)$ instead.

Similarly, the *two-sided prediction interval* at the $p \cdot 100$ *percentile* is constructed as:

$$\phi_t^2(x_t, p) = \widehat{\mu}(x_t) \pm |\, q((1-p)/2)\,| \cdot \widehat{\sigma}(X_t'). \tag{5.5}$$

Whenever the context is clear, we omit x_t and refer to the right endpoint of the interval as $\phi_t^{2\triangle}(p)$ and to the left endpoint as $\phi_t^{2\nabla}(p)$, respectively.

5.2.2 Evaluation Metrics

Practitioners, e.g., trading power demand/supply in day-ahead electricity markets, require reliable prediction intervals. A key performance indicator is: *Does the prediction interval for the 95 percentile, in the long run, cover the demand indeed 95% of the time?*[4] Additionally, smaller interval sizes are preferred over too conservative uncertainty estimates, since they generally lead to more efficient operational decisions. The pinball loss function is widely used in the literature to assess the accuracy of probabilistic/quantile forecast [Koenker and Bassett, 1978, Takeuchi et al., 2006]. However, it does not express how reliable or how wide the prediction intervals are, and thus provides only limited insights into the quality of uncertainty forecasts. To this end, we propose evaluation metrics that measure explicitly the quantities of interest above, i.e., the coverage accuracy and the width of the intervals.

Empirical coverage We define the empirical one-sided coverage of our estimated intervals for the $p \cdot 100$ percentile during a time period $T = \{t_1, \ldots, t_{|T|}\}$ as:

$$c_1(p, T) = \frac{1}{|T|} \sum_{t \in T} \mathbf{1}_{(y_t \leq \phi_t^1(p))}. \tag{5.6}$$

This can be seen as an estimate of $Pr(Y_t \leq \phi_t^1(p))$ over a time period T. The closer the coverage is to p, the better. Similarly, the empirical two-sided coverage is defined as:

$$c_2(p, T) = \frac{1}{|T|} \sum_{t \in T} \mathbf{1}_{(\phi_t^{2\nabla}(p) \leq y_t \leq \phi_t^{2\triangle}(p))}. \tag{5.7}$$

Similar to the one-sided case, we can see this is an estimate of $Pr(\phi_t^{2\nabla}(p) \leq Y_t \leq \phi_t^{2\triangle}(p))$ over a time period T, and the closer this coverage is to p, the better.

Coverage absolute error (CAE) Given the estimates of one- and two-sided coverages for the $p \cdot 100$ percentile during a time period T, we define the CAE of the one-sided interval as $|p - c_1(p, T)|$ and the CAE of the two-sided interval as $|p - c_2(p, T)|$.

Mean percentage width (MPW) We define the MPW of the one-sided interval at the $p \cdot 100$ percentile as: $\frac{1}{|T|} \sum_{t \in T} |\phi_t^1(x_t, p) - \mu(x_t)|$, and the MPW of the two sided interval as: $\frac{1}{2 \cdot |T|} \sum_{t \in T} \left(\phi_t^{2\triangle}(p) - \phi_t^{2\nabla}(p)\right)$. Note that, we divide the latter by two so that it has the same dimension as the one-sided version.

[4] See also [Weron, 2014, Wytock and Kolter, 2013]. Note that, reliability over other percentiles can be expressed similarly, we use 95 percentile only as an example.

5.3 Online Learning

In order to track trends in the mean demand and keep the forecasting models up to date, we leveraged the online learning algorithm outlined by Ba et al. [2012]. In this section, however, we outline another, complementary online learning algorithm to adapt the prediction intervals to the non-stationary nature of the electricity demand. That is, when the prediction interval achieves the desired coverage, then the adaptive construction algorithm practically has no effect. On the other hand, when the prediction interval starts to cover lower or higher percentile than expected, the algorithm adjust the interval by gradually widening or narrowing it.

Let us denote the empirical coverage as described in Eq. 5.6 and 5.7 as c. Then, the coverage $\widehat{c}$ that we should target for the next time periods (or construction horizon) h to cover the desired p portion of the data (or the $p \cdot 100$ percentile) is given by:

$$\begin{aligned} \alpha\ c + (1-\alpha)\widehat{c} &= p \\ \widehat{c} &= \frac{p - \alpha\ c}{1-\alpha}, \end{aligned} \tag{5.8}$$

where we call $\alpha \in [0, 1)$ the *aggressivity* parameter. The higher the aggressivity parameter α, the more eager we are in adjusting $\widehat{c}$ to cover the $p \cdot 100$ percentile. The construction horizon h should be equal to or greater than the forecasting horizon. Additionally, since Eq. 5.8 might yield values smaller than 0 or greater than 1, we set the minimum value of $\widehat{c}$ to ε, and the maximum value to $1 - \varepsilon$, where ε is a positive number close to zero.[5]

Let q be the quantile function of the standard normal distribution. Then, for the next time periods h, we construct the (adaptive) intervals as follows.

Adaptive one-sided interval We define the adaptive one-sided interval as

$$\phi_t^1(x_t, p) = \widehat{\mu}(x_t) + q(\widehat{c}) \cdot \widehat{\sigma}(x_t), \tag{5.9}$$

i.e., we modify the one-sided interval construction in Eq. 5.4 by using $q(\widehat{c})$ instead of $q(p)$.

Adaptive two-sided interval Similarly, we define the adaptive two-sided interval to cover p portion of the data as

$$\phi_t^2(x_t, p) = \widehat{\mu}(X_t) \pm |\, q((1-\widehat{c})/2)\,| \cdot \widehat{\sigma}(x_t), \tag{5.10}$$

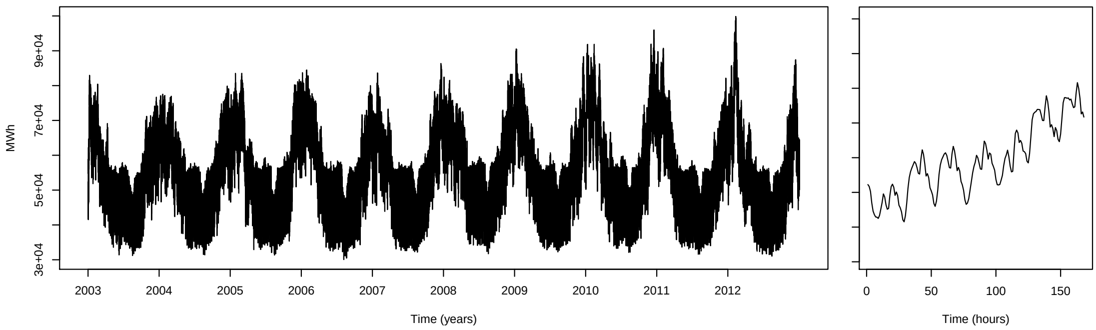
i.e., we modify the two-sided interval construction in Eq. 5.5 by using $q((1-\widehat{c})/2)$ instead of $q((1-p)/2)$.

In the implementation, we assume that every day at midnight (one could also choose a different time), we provide demand forecasts with uncertainty for the next 24 hours (i.e., $h = 24$ hours). After the 24 hours have elapsed, we use the observed actual demand values to update our model.

5.4 Experimental Results

We focus on day-ahead forecasts, where the demand forecast and its prediction interval for each hour for the next 24 hours are delivered once per day at a specific time of day (e.g., midnight).

[5] In our experiments, we use $\varepsilon = 0.0001$.



(a) (b)

Figure 5.1: Hourly electricity demand in France; (a) the complete view of the dataset (from January 2003 to December 2012), and (b) the first week of the dataset.

Table 5.1: Our out-of-sample, non-overlapping, rolling window test period.

Evaluation	Train period	Test period
#1	Jan 2003 - Dec 2010	Jan 2011 - Dec 2011
#2	Jan 2004 - Dec 2011	Jan 2012 - Dec 2012

5.4.1 Dataset

We use the publicly available dataset provided by the Réseau de transport d'électricité (or simply RTE dataset)[6] which contains the national electricity demand of France from January 2003 to December 2012. While the original data is recorded every 30 minutes, we aggregated the measurement into hourly measurements. The resulting time series is shown in Figure 5.1: in (a) one can see clear seasonal cycles over the 10 years, while (b) shows typical daily patterns of electricity demand. Notice the overall increasing trend in demand and the unusually high peak in winter 2011/2012 that are interesting challenges for evaluating the effectiveness of our online learning mechanisms.

5.4.2 Forecasting Uncertainty

We evaluate the accuracy of our model using out-of-sample, non-overlapping, rolling window test periods. There are two evaluation windows (see Table 5.1). We use the first eight years as the train period and the subsequent year as the test period.

[6]See http://clients.rte-france.com/lang/an/visiteurs/vie/vie_stats_conso_inst.jsp, accessed: 2014-02-28. Additionally, we have made our curated dataset (including holidays and weather information) available at https://github.com/tritritri/uncertainty.

Modeling the conditional mean In this experiment, we model the conditional mean of the electricity demand using the following GAM:

$$\begin{aligned}\log(\mathbb{E}[Y_t]) = \beta_0 + \beta_1 \texttt{tmrwDayType}_t + \\ \textstyle\sum_{j=1}^{8} \mathbf{1}_{(\texttt{DayType}_t = j)} f_{1,j}(\texttt{HourOfDay}_t) + \\ f_2(\texttt{TimeOfYear}_t) + f_3(\texttt{TempC}_t, \texttt{Humidity}_t) + \\ f_4(\texttt{TempC}_t, \texttt{HourOfDay}_t) + f_5(\texttt{LagNSameDTLoad}_t) + \\ \textstyle\sum_{j=1}^{8} \mathbf{1}_{(\texttt{DayType}_t = j)} f_{6,j}(\texttt{LagLoad}_t),\end{aligned} \tag{5.11}$$

The covariate `DayType` is a categorical variable with 8 values that represent the seven days of the week (from Sunday to Saturday) and public holidays, `tmrwDayType` models transitions between different day types,[7] `TimeOfYear` models yearly seasonality,[8] `LagLoad` is the load for the same hour of the previous day, and `LagNSameDTLoad` is the average load for the same hour of the previous two days of the same categories (i.e., weekday/weekend). The covariate `TempC` (and `Humidity`) are the weighted average of the temperature (and the humidity) of the 6 biggest cities in France, proportional to their annual energy consumption. We illustrate some transfer functions in Figure 5.2. The model achieves a Mean Absolute Percentage Error (MAPE) of 1.65 (averaged over the two test periods).

Modeling the conditional variance As described in Section 5.2.1, we obtain the estimated conditional variance function $\widehat{\sigma}^2(\cdot)$ by fitting a second GAM to the squared empirical residuals $\widehat{z}_t^2$. We model it as follows:

$$\begin{aligned}\log(\mathbb{E}[\widehat{z}_t^2]) = \beta_0 + \beta_1 \texttt{DayType}_t + \\ \textstyle\sum_{j=1}^{5} \mathbf{1}_{(\texttt{DayType}_t = j)} f_{1,j}(\texttt{HourOfDay}_t) + \\ f_2(\texttt{TimeOfYear}_t) + f_3(\texttt{TempC}_\texttt{t}, \texttt{Humidity}_t) + \\ f_4(\texttt{TempC}_t, \texttt{HourOfDay}_t).\end{aligned} \tag{5.12}$$

To justify our normality assumption and evaluate the effectiveness of our model above, we show the normal Q-Q plot of the empirical residuals $\widehat{z}_t$ (Figure 5.3a) and the rescaled empirical residuals $\widehat{\epsilon}_t$ (Figure 5.3b). The rescaled residuals are indeed much more normal than the original residuals $\widehat{z}_t$. Figure 5.4 illustrates the prediction intervals for the first week of the test set #1 (see Table 5.1). Next, to evaluate the accuracy and the width of the prediction intervals we compute their CAE and MPW. Figure 5.5 shows that the coverage error (CAE) is around 1%-2% (approximately at the same level as the forecasting error for the conditional mean) with lower error and higher interval width (MPW) in the tail. Moreover, note that our MPW are fairly small (mostly around 3%-4%). For instance, if the MPW at the 95 percentile is around 4% and the average demand is around 55000 MWh, then on average, the two-sided prediction interval for the 95 percentile is around $[52800, 57200]$ MWh.[9]

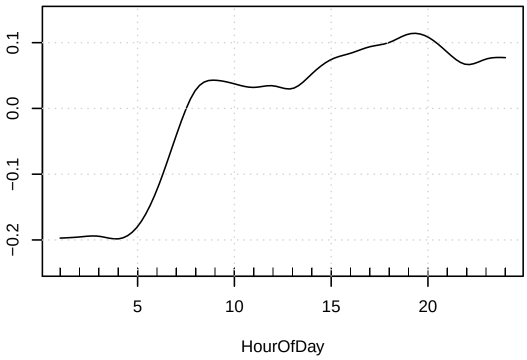
5.4.3 Online Learning

Although we have already relatively small prediction errors (1%-2%), this high accuracy might not be sustained over a long period time. In a real-world implementation, a drop in model performance could be caused by trends and dynamic changes in the way people consume electricity,

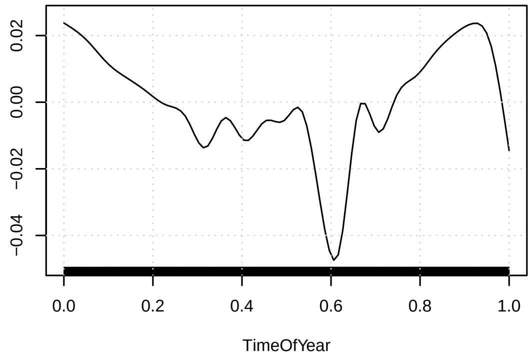
[7] It is a 2-gram consists of today's and tomorrow's day type.

[8] It is an increasing value from 0 to 1, repeated yearly from the beginning to the end of the year.

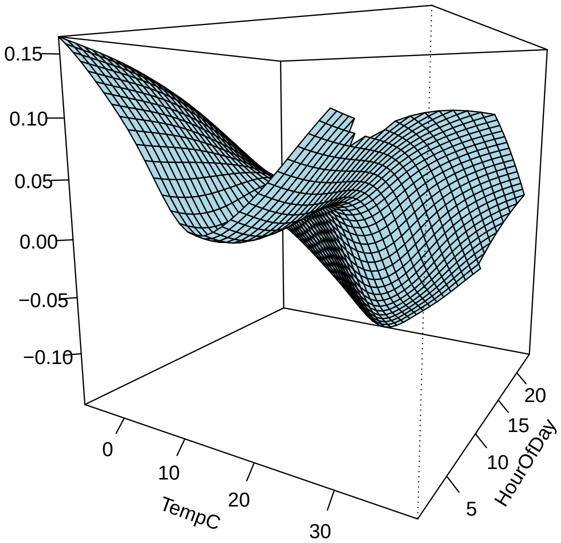
[9] The demand in Figure 5.4 is around 75000 MWh since it shows the demand in the beginning of January (i.e., during the winter season), which is typically higher than yearly average in France.



(a) $f_{1,2}$(`HourOfDay`)



(b) f_2(`TimeOfYear`)



(c) f_4(`TempC`, `HourOfDay`)

Figure 5.2: Visualization of some transfer functions from the model in Eq. 5.11. It illustrates an additional benefit of using GAM, i.e., its transfer functions provide insights about the relationship between explanatory and target variables: (a) represents the typical hourly load curve on Monday, (b) the yearly seasonality of electricity demand, where the winter demand is typically lower than that of summer, (c) the joint effect of temperature and hour of the day. The dip around `TimeOfYear` $= 0.6$ in (b) shows the effect of school summer holiday. Note that, the function outputs above are of small magnitude since the model considers the logarithmic of the demand.

e.g., related to new electricity tariffs, the emergence of new appliances, or changes in macroeconomic conditions. In this section, we show the effectiveness of the online learning mechanism in addressing these challenges and maintaining high forecasting accuracy. For this experiment, we use the first two years of the RTE data set as the train period and the last eight years as the test period.

Figure 5.6 shows the forecasting error (MAPE) of both models, without and with online learning (described in [Ba et al., 2012]). Although at the beginning both models have similar performance, the MAPE of the model without online learning continues to increase over time,

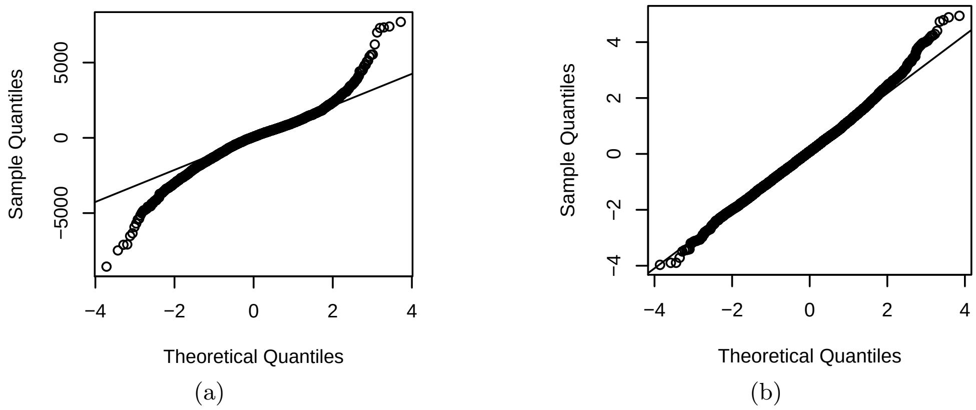


Figure 5.3: Normal Q-Q plot of (a) the empirical residuals $\widehat{z}_t$, (b) the rescaled residuals $\widehat{\epsilon}_t$.

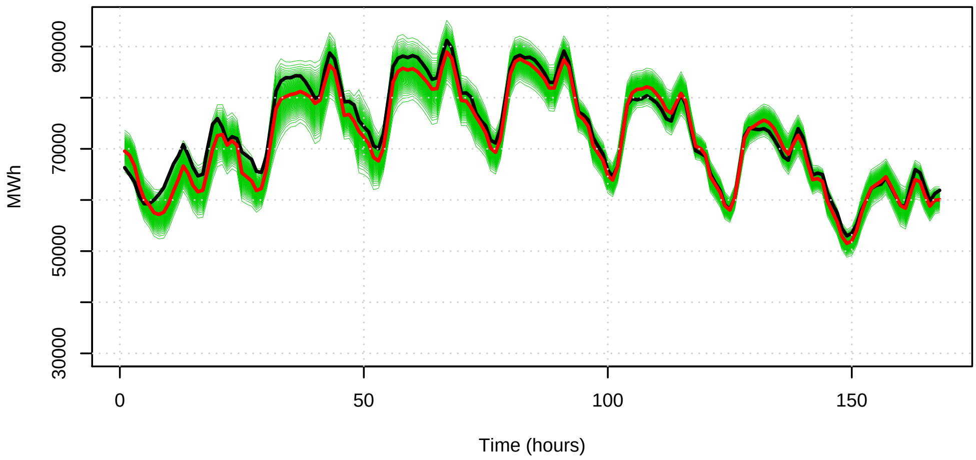
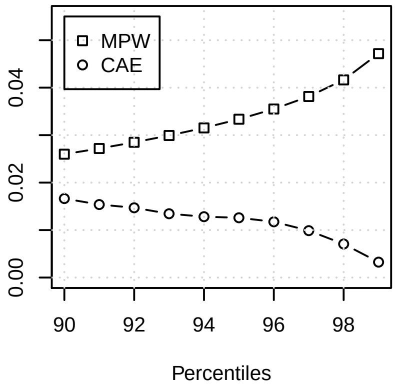


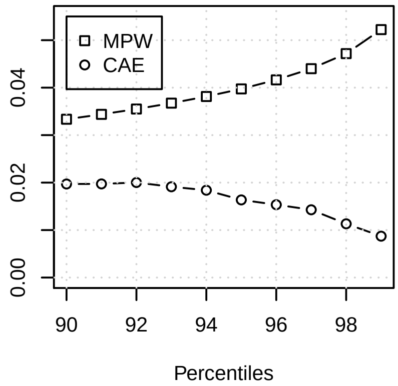
Figure 5.4: Prediction intervals of the first week (Sunday–Saturday) of the test set #1 (see Table 5.1). The black line is the actual demand, the red line is the forecasted expected demand, and the green lines are the prediction intervals (from bottom to top: one-sided prediction intervals for 1 to 99 percentiles).

whereas with online learning we are able to keep the error low. Intuitively, using the online learning mechanism, the model is able to adapt to the dynamic changes of the electricity demand, and thus prevent the MAPE from increasing.

Next, to evaluate the effectiveness of the adaptive construction of the prediction intervals (described in Section 5.3), we compute the coverage of the intervals without and with the adaptive construction for various percentiles, i.e., 90, 80, ..., 20, 10 percentiles. For the adaptive construction, we use the parameters $\alpha = 0.95$ and $h = 24$ hours (equal to the forecasting horizon). Figure 5.7 shows that the adaptive construction successfully maintains the coverage accuracy for all the percentiles shown, whereas without it, the accuracy tends to decrease over time (with lower accuracy for the higher percentiles).



(a) one-sided intervals



(b) two-sided intervals

Figure 5.5: The mean percentage width (MPW) and the coverage absolute error (CAE) of the estimated (a) one- and (b) two sided prediction intervals, averaged over the two test periods.

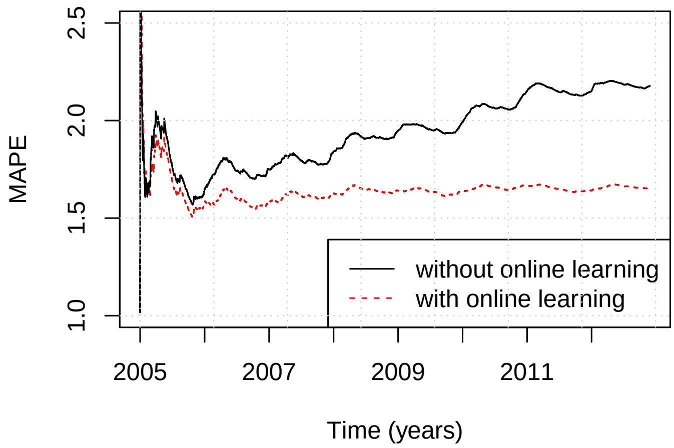


Figure 5.6: The forecasting error (MAPE) over time for the two models, i.e., with and without online learning. The online learning mechanism succeeds to keep the forecasting error low over time.

5.5 Summary and Discussion

This chapter presents a novel GAM^2 algorithm to forecast uncertainty in electricity demand by modeling the time-varying conditional mean and variance. Our method is efficient since it does not require multiple runs of simulations or bootstrapping. We assess the coverage accuracy as well as the width of the prediction intervals. Furthermore, we incorporate online learning mechanisms to adapt the forecasted mean and prediction intervals to the dynamic changes in the distribution of the demand. Although we showcase our method specifically for electricity demand, it can also be applied to other domain as well, as long as the rescaled residual is approximately normal.

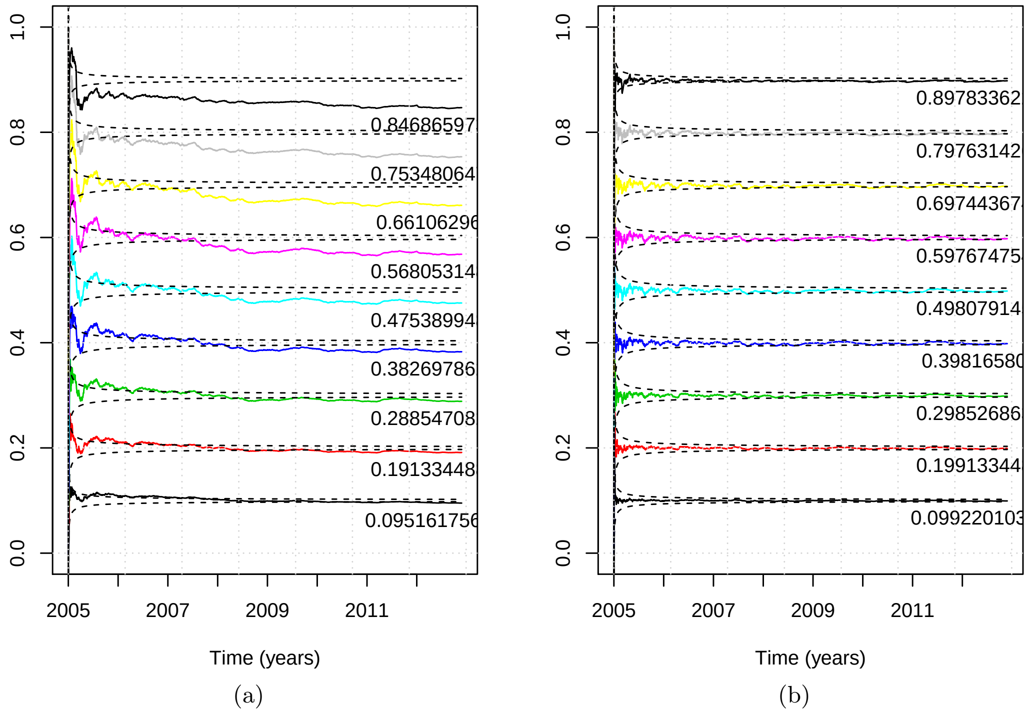


Figure 5.7: Empirical coverage (a) without and (b) with adaptive interval construction over time. The solid lines represent the two-sided coverages for the (from top to bottom) 90, 80, ..., 20, and 10 percentiles. The numbers on the right hand side show the empirical coverages at the end of the test period. The dotted lines are the binomial (95%) confidence interval for each percentile, which serve as a guide on approximate ranges of acceptable coverages.

Reducing Data Size and Privacy Risk 6

While smart meter data analytics allow utility companies to analyze customer consumption behavior in real time, the amount of data generated by these sensors is very large. As a result, analytics performed on top of it typically require expensive storage and computational cost. Furthermore, smart meter data contains very detailed energy consumption measurement that could lead to customer privacy breach and all risks associated with it. This addresses the problem on how to reduce smart meter data numerosity and its detailed measurement while maintaining its analytics accuracy. We convert the data into symbolic representation and allow various machine learning algorithms to be performed on top of it. In addition, our symbolic representation admit an additional advantage to allow also algorithms which usually work on nominal and string to be run on top of smart meter data. We provide an experiment for classification and forecasting tasks using real-world data.

This chapter has also appeared in Proceedings of the Joint EDBT/ICDT 2013 Workshops [see Wijaya et al., 2013b].

6.1 Introduction

Smart meters are widely deployed for various reasons, such as making people aware of their consumption, understanding the usage pattern of appliances , or enabling fine grained energy billing. They also contribute to the creation of "Big Data" and participate in the "Internet of Things." But this trend of digitalizing every single expression of our surrounding world has also some drawbacks. First, the data collected could contain personally identifiable informations that people may not want to reveal or make accessible. This is also the case with smart meters where customer habits can be extracted from highly detailed power consumption data [Albert and Rajagopal, 2013, Beckel et al., 2013, Weiss et al., 2012]. Secondly, for utility companies, managing a huge size of smart meter data and performing analytics on top of it requires more and more resources (time, energy, machines).

However, only small parts of this Big Data that might be relevant and contain the few information bits that are needed for the analytics at utility companies. Therefore extracting this information before storing the data or even before centralizing it may help reducing its overall size and could also be used to hide private parts.

One approach is to develop compression that can be applied at the sensor level, well known as compressive sensing methods [Baraniuk, 2007], which are designed such that the sparsity of the information could help reconstruct the real values from the compressed ones. In the case of smart meters one can think about the information as the user switching devices on and off, this is very sparse compared to the amount of real values generated every second by the smart meters. But as our goal is not to reconstruct perfectly the real values, but to perform analytics on the compressed ones, we explore another compact representation, namely symbolic representation.

Overview of Contributions This presents our approach for replacing large time series of smart meter data with smaller sequences of symbols that can still support statistics and machine learning algorithms for some selected purposes, such as customer segmentation or consumption forecasting. We summarize our contributions as follows.

- We propose symbolic representation methods for time series which admit online conversion (Section 6.2).
- We show how to apply our symbolic representation for two analytics tasks in the energy domain (which can be generalized), namely customer segmentation and consumption forecasting and evaluate them using a real-world dataset (Section 6.3).

6.2 Symbolic Representation

The data generated by sensors and in particular smart meters is represented as time series of measurements at the rate given by the specific sensor. In this section, we give the formal definition of our data (time series), and describe our approach to convert the raw data into symbols.

Definition 6.1 (Time Series)**.** *We define time series formally as* $S = \{s_1, s_2, s_3, \ldots\}$*. Each* $s_i \in S$ *is a two tuple* (t_i, v_i)*, where* t_i *is a timestamp and* $v_i \in \mathbb{R}$ *is a measurement on time* t_i*. Whenever* $j \leq i$*, we require* t_i *to be no earlier than* t_j*.*

Our symbolic representation aims at representing a contiguous portion of a time series as a single symbol. Symbols are taken from a predefined alphabet and in our case we will consider binary numbers, such as '0','101' or '00101'. As we can consider different symbols of different size, our alphabet will only have a partial order, '0' being equal to '01', '00' and so on. To allow reconstruction of the original time series, we also build a lookup table that will match each symbol to the average real value of it corresponding range. More specifically in the case of smart meters, the lookup table is built once at the sensor level and then sent to the aggregation server before starting to send the symbolic data. We can also consider rebuilding and resending the lookup table periodically or if the distribution of the data changes too much.

Symbolic representation needs two kinds of aggregations and therefore two kinds of segmentations.

6.2.1 Vertical Segmentation

Vertical segmentation is a temporal aggregation over real valued time series. It aims to reduce the numerosity of the data. In principle, we can use any aggregation method, such as average, sum, maximum/minimum value, etc. In our approach, we use average.

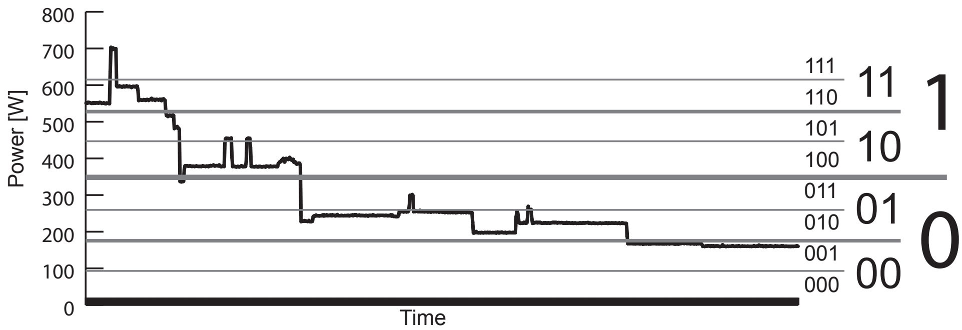


Figure 6.1: Construction of the variable length of symbols by recursive division of the real values range.

Definition 6.2 (Average Vertical Segmentation). *Given time series* $S = \{s_1, s_2, s_3, \ldots\}$*, then* $\mathcal{V}_A(S, n) = \{\bar{s}_1, \bar{s}_2, \bar{s}_3, \ldots\}$ *is a new time series resulting from vertical segmentation of* S *which averages* n *values together. Each* $\bar{s}_i = (\bar{t}_i, \bar{v}_i)$*, where* $\bar{t}_i = t_{i*n}$*, and*

$$\bar{v}_i = \frac{1}{n} \sum_{(i-1)*n+1}^{i*n} v_i.$$

6.2.2 Horizontal Segmentation

Horizontal segmentation changes the granularity of the values a symbol can represent. Depending on the requirement, several strategies are conceivable. To allow variability in the granularity of the symbols, we build our binary symbols of variable length by recursively dividing the sub-ranges of real values, as shown on Figure 6.1.

Definition 6.3 (Horizontal Segmentation). *We define a* table lookup *as two tuple* $L = (\mathcal{A}, B)$*, where* $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ *is our alphabet (a set of symbols) and* B *as a set of separators* $\beta_i \in \mathbb{R}$ *where* $1 \leq i \leq k-1$*.*

Given time series $S = \{s_1, s_2, s_3, \ldots\}$*, then* $\mathcal{H}(S, L) = \{\hat{s}_1, \hat{s}_2, \hat{s}_3, \ldots\}$ *is a symbolic time series resulting from horizontal segmentation of* S *using table lookup* L*. Each* $\hat{s}_i = (t_i, \hat{v}_i)$*, where* $\hat{v}_i \in \mathcal{A}$*, and:*

(i) if $v_i \leq \beta_1$*, then* $\hat{v}_i = a_1$*,*

(ii) if $v_i > \beta_{k-1}$*, then* $\hat{v}_i = a_k$*, otherwise*

(iii) $\hat{v}_i = a_j$*, where* $\beta_{j-1} < v_i \leq \beta_j$*.*

From Definition 6.3, separators and lookup tables play an important role to determine the end result of the horizontal segmentation. We learn the *separators* using information from data distribution. Additionally, we propose three different methods to generate such lookup tables, namely *uniform*, *median* and *median of distinct values*. For the division process, we propose the following methods:

a) *uniform*: each symbol is assigned ranges of values of same size. To compute them we first get the maximum value, *max*, and divide uniformly the range from zero to *max* in k subranges

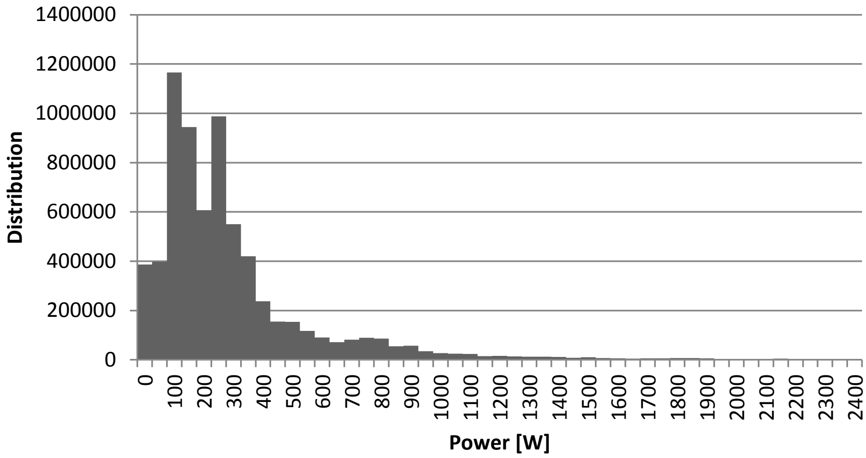


Figure 6.2: Distribution of Energy levels with 1 second sampling rate in the REDD dataset follows a log-normal distribution.

for each of the k symbols. Hence, $\beta_1 = \frac{max}{k}, max - \beta_{k-1} = \frac{max}{k}$, and $\beta_i - \beta_j = \frac{max}{k}$ for $i = j + 1$.

b) *median*: each symbol represents the same number of real values. We divide the ordered data into k equal-sized data subsets (*k-quantiles*). For the separators, we take the data values as the boundary of consecutive subsets, i.e. β_1 is the the data value in the boundary between the first and the second subset, β_2 is the boundary between the second and the third subset, and so on. This horizontal segmentation aims to maximize the entropy of the generated symbols.

c) *median of distinct values*: each symbol represents the same number of distinct real values. We take all measurement values in our time series as a set. Then, we order the elements in the set. Next, we use k-quantiles to determine the separators (similar to *median* start from this step). If the real values have enough precision to always be different this becomes equivalent to *median*, however if some value v occurs very often, this method avoid a bias toward v. We also refer to this method as *distinctmedian* for short.

Of course if the overall distribution of the real values is perfectly uniform and limited to a fixed range, these three methods are equivalent.

The most similar approaches in the literature so far are SAX [Lin et al., 2007] and iSAX [Shieh and Keogh, 2008]. But they were designed to run offline with a fixed alphabet size, $\mathcal{A}$ with a fixed k, and assuming that the distribution of v_i is Gaussian. Using this assumption, the v_i are first normalized and then separators β_i are taken at pre-defined values from a table such that they divide equally the samples. In our case the distribution is log-normal as shown on the Figure 6.2 and our *median* approach is a generalization of the method used for SAX. Moreover, individual normalization per house would not allow us to differentiate big customers from the small ones as we can see in Figure 6.3, where customer A and B would be normalized to the same level as C and D respectively. Therefore we need to develop other methods for generating the symbols.

On real-time systems, data is continuously produced and processing can be done either online each time a new value is measured, or by batches on past values. But, of course the algorithm

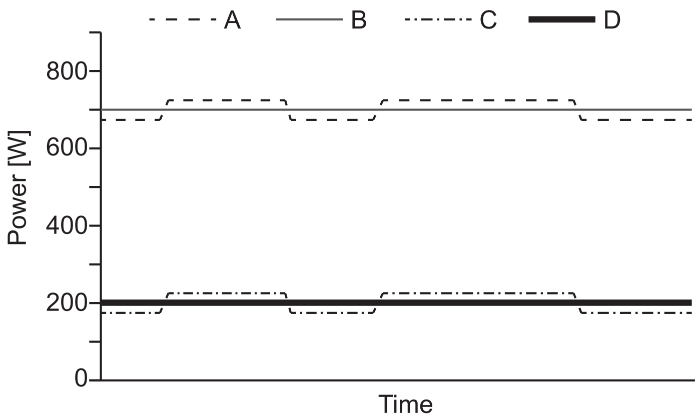


Figure 6.3: Without normalization A and B (C and D) are more similar, but with normalization A and C (resp. B and D) would be put together.

cannot use future data. In our conversion process to symbols, the first horizontal segmentation has to be performed before the system can start to process any data. Indeed, the range of values assigned to a symbol cannot change too often, otherwise it would need to propagate also the new lookup table and make difficult to run any algorithm on top of the symbols as they may have different representations. Therefore, we propose to use historical data for determining the distribution of the real values and compute the separators according to the three methods described above. The time period used should be representative for the typical behavior of the measured phenomenon. In our case of power consumption measurement, it should include day and night or weekday and weekends values.

For horizontal segmentation, there exist online algorithms which are able to select dynamically when a new segment has to be started [Fuchs et al., 2010, Guo et al., 2012, Liu et al., 2008]. But in our case, variation of the time dimension of a symbol would give it a different semantics and thus preventing us to run algorithms on top of symbolic data. For this reason we chose to predefine the length of the temporal aggregation and evaluate different options.

6.2.3 Compression Ratio

The compression ratio depends on both the vertical and the horizontal segmentation granularity. For example, if the original data is stored as *double* (64bit) and sampled at 1 Hz, we have around 680 kB of data per day. Now if we use 16 symbols and an aggregation of 15 minutes, it would leave us with only 384 bit, three order of magnitude lower. Of course the lookup table should be taken into account, but it can be amortized over time. Communication and storage overhead, induced for example by protocols and indexes should also be taken into account for a real system.

6.3 Experiments

There are three smart meter datasets available which are based on real measurements:

- REDD dataset [Kolter and Johnson, 2011]: contains data of 6 houses for appliance and house level consumption. The measurement is done every second for the duration of 1 to 2 months.

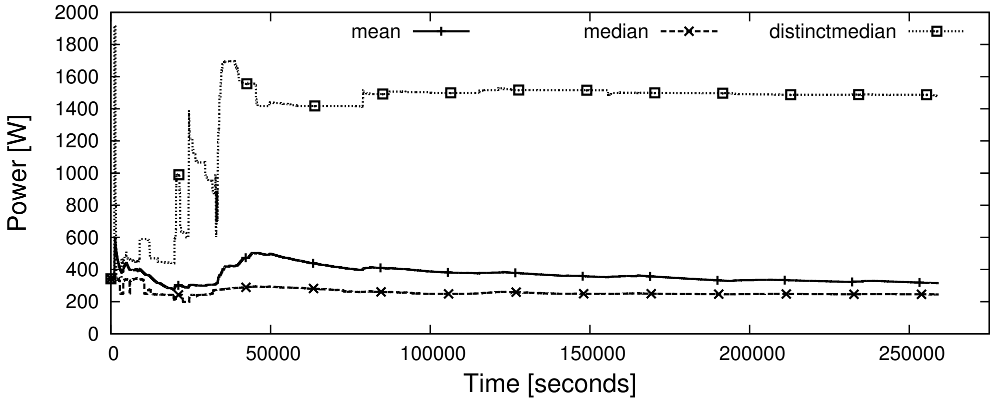


Figure 6.4: Statistics on energy consumption of house 1 in REDD dataset (86400 seconds = 1 day).

- Smart* dataset [Barker et al., 2012]: contains data of 443 houses for house level consumption for the duration of 24 hours. In addition, it has three houses with more fine grained house level measurements (every second) for around 3 months, where one of them also has circuit level measurement.
- Irish CER dataset [The Commission for Energy Regulation (CER) and the ISSDA, 2012]: contains data of around 5000 houses for house-level consumption. The measurement is done every 30 minutes for the duration of one and a half years (see also Section 3.5.1).

For our experiments, we decided to use REDD dataset as a tradeoff between duration and granularity. It has more fine grained measurements compared to Irish CER dataset, and it has a longer measurement duration compared to Smart* dataset.

For each of the separators generation methods presented in the previous section, we used the first two days of data for each house to compute the necessary statistics. This allows us to get an estimation of the distribution of the data, without using the complete dataset. As an illustration, Figure 6.4 shows the accumulative mean, median, and median of distinct values of house 1 in REDD dataset for three consecutive days. The statistics start to converge after day one.

Our symbolic encoding procedure needs two parameters, namely the temporal aggregation length and the size of the symbols alphabet.

Aggregation length The aggregation length determines the smoothing of the data and can be seen as a low-pass filter. It can also be applied directly on the raw values to highlight the effects of the other parameter only. We choose 15 minutes and 1 hour as they reflect the typical segmentation in smart energy algorithms (see, e.g., [Mohsenian-Rad et al., 2010, Wijaya et al., 2013c]).

Alphabet size The size of the alphabet determines the granularity of the data and the accuracy of the reconstruction. However, too many symbol variations would make us loose the advantages of efficiency in data processing and machine learning. Furthermore, as our symbols are stored as binary numbers, we used only the power of 2. In particular, we look at different values between 2 and 16.

We used the total power consumption of the house, by summing the two main power time series for each house. In the following, we look at two main scenarios of energy data management: customer segmentation and consumption forecasting.

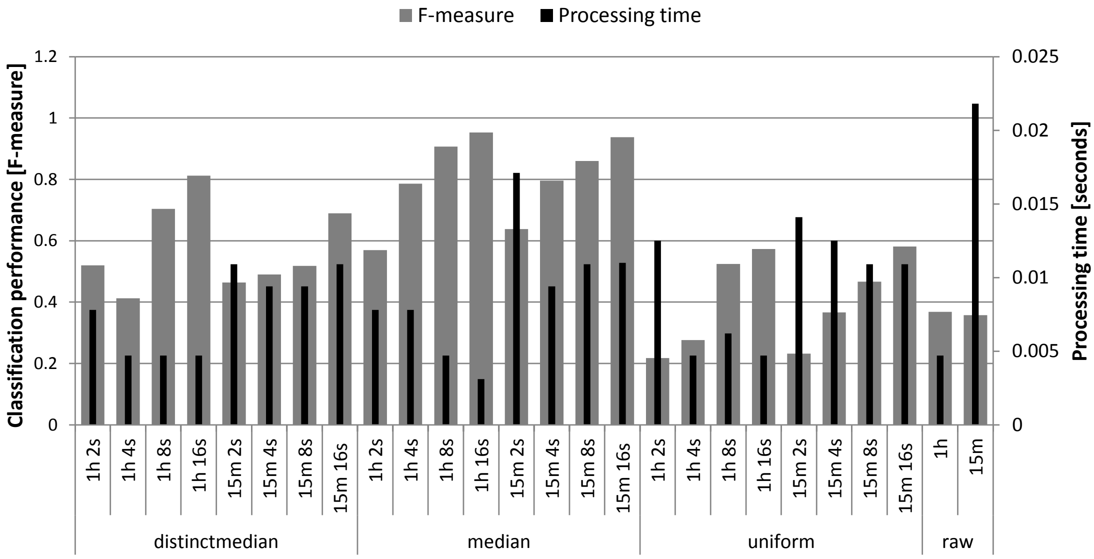


Figure 6.5: Evaluation of a Naive Bayes classifier over symbolic and raw data.

6.3.1 Classification

Identifying customers having a similar consumption profile (customer segmentation) is very relevant for the future of smart grid, especially for demand response system and intelligent distribution channel. However, as we only have 6 houses in our dataset, we consider each house having its own cluster and split it by days. Then, we take some of these days for training and the rest for evaluating the classification. This follows the same assumption as for the previously presented distribution estimation, that users behave similarly over time. Interestingly this could also be seen as an attack against changing ID's privacy protection mechanisms such as the one used in [Bindschaedler et al., 2012] that could be also applied for anonymizing smart meters data.

As the dataset contains gaps (missing values), we select days where the house has *enough* data, putting the threshold at 20h per day of data and build one vector per day, the class label being the house number. The so generated files were used as input for Weka's [Hall et al., 2009] implementation of various classifiers such as Random Forest, Logistic, J48, and Naive Bayes. Another advantages of our approach is that it is not linked to any specific classifier. Hence, all algorithms supporting nominal values can be applied.

Evaluation using 10-fold cross-validation gives the results shown on Figures 6.5 and 6.6. The timing value was computed as the average over 10 runs and the classification performances are evaluated with the F-measure (the weighted harmonic mean of Precision and Recall). To have vectors of same size, raw values were also aggregated, by taking the average over 15 minutes, respectively 1 hour.

Time-wise, 15 minutes aggregation is in average slower than the corresponding 1 hour ones and the raw dataset always took slightly longer to process, mostly because it was composed of numerical values instead of symbols. The running time over the full raw vectors (actual measurements, without aggregation) was much slower by two orders of magnitude. Hence, we choose not to include it in the figures.

As expected, the classification performances vary depending on the algorithm used, but it

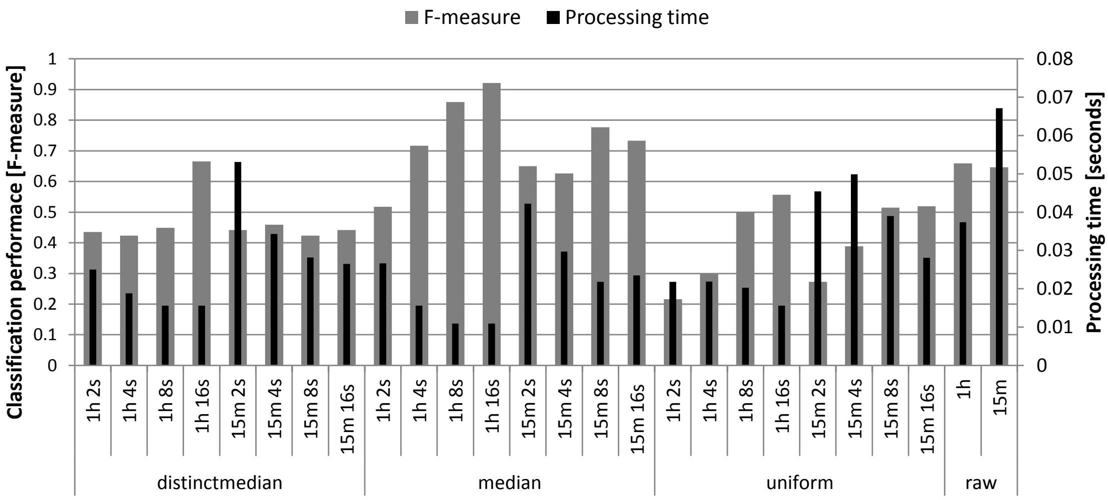


Figure 6.6: Evaluation of a Random Forest classifier over symbolic and raw data.

is consistent across the different parameters. Therefore, we show only the figures for Naive Bayes and Random Forest. Accuracy improves with the size of the alphabet as more detailed information, with a better granularity, are kept during the conversion process. On the contrary, one hour aggregate performs sometimes better than the 15 minutes ones, especially with 16 symbols. This can be explained as the noise filtering effect of our low-pass filter.

The classification using raw values, however, depends much more on the algorithm and Random Forest is the one performing better. But even with this algorithm, it is not able to outperform *median* encoding performance. To check if this is due to a loss of information by aggregating values, we run also the Random Forest classifier over the full dataset by using the same day separation. However, there is only an accuracy increase of 4%. This is not enough to catch the performance of the *median* encoding.

On average, *median* encoding performs better than *distinctmedian*, which is better than *uniform*. Indeed, by choosing to maximize the representativity of a symbol in the *median* scheme, it allows classifiers to better differentiate the houses, as their encoding is based on their own statistics, adding information to the encoded data. The choice of the delimiters containing information specific to the house (quantiles for example), the encoding will generate sequences of symbols that are typical to the considered house, thus making them more differentiable and classifiable.

To verify the hypothesis above, we ran again the encoding, but by using statistics over all houses instead. The results are show in Figure 6.7. Overall, the performance of classification on symbolic data is decreased. However, using Random Forest classifier, *median* encoding still manage to reach the same level as the raw values. And, using Naive Bayes classifier it even outperformed it (see Table 6.1). This shows that this encoding preserves all needed information for performing classification over symbolic representation of sensor data. Table 6.1 summarizes all our classification experiments.

6.3.2 Forecasting

Consumption forecasting is an important task in the smart grid for determining what action utility companies should take given the energy availability. In this experiment, we aim to do

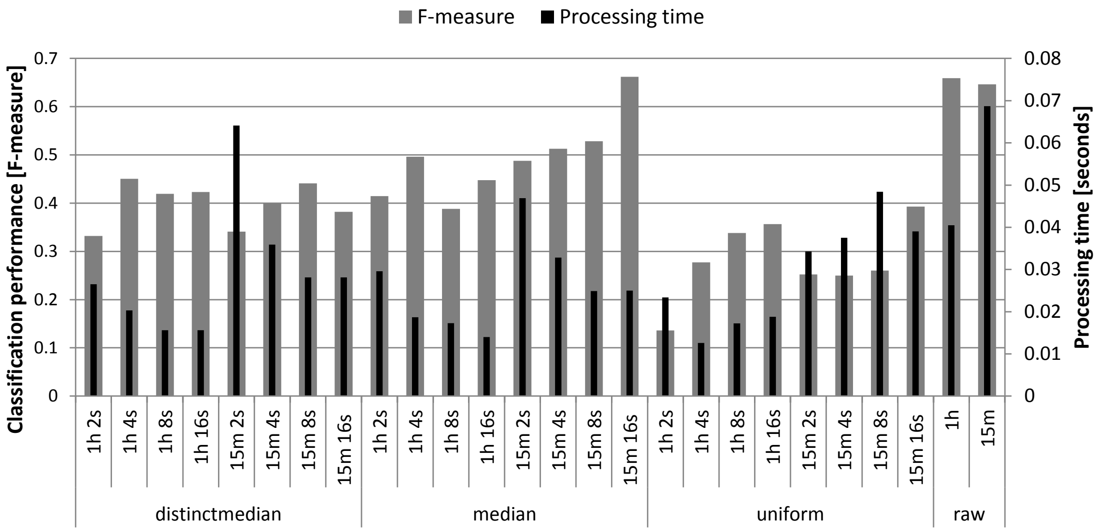


Figure 6.7: Evaluation of a Random Forest classifier over symbolic (using a single lookup table) and raw data.

short-term hourly residential load forecasting. For residential consumption, even in the same activity, a small change in the order of which appliances are turned on results in totally different pattern. While consumption pattern for a large, industrial scale is more plausible and has been widely studied [Charytoniuk et al., 1998, Chen et al., 2004, Hippert et al., 2001, Papalexopoulos and Hesterberg, 1990, Sapankevych and Sankar, 2009, Taylor, 2010], this is not the case with consumption pattern for an individual house.

We show a case where we predict the next day electricity consumption of a house given 1 week historical data. When we do forecasting using symbols, it means that we forecast the next symbols. In our symbolic representation, each symbol represents a range of values. However, in the consumption forecasting task, what we need are the real consumption values (instead of symbols). For this matter, in our experiment, we define semantics of a symbol as the center of its range. At first, it seems that this exposes the disadvantage of using symbols in forecasting. But, as we can see later, performance of our symbolic forecasting is comparable with (in some cases better than) forecasting using the real values.

For each house, we take 1 week hourly consumption data as training and the next day hourly consumption data for testing. We perform symbolic forecasting using *median*, *distictmedian*, and *uniform* using alphabet of length 16. We choose this length as a trade off between symbol range's granularity and encoding length. We reduce the forecasting task into classification task using lag attributes of length 12 comprises of 12 previous symbols. The target attribute is the next symbols. For the algorithms, in principle we can use any machine learning algorithm for classification. However, in this experiment we show the result using Random Forest and Naive Bayes.

As a comparison to our symbolic forecasting performance, we also perform consumption forecasting on real values. Several techniques have been suggested in the literature, e.g., regression [Charytoniuk et al., 1998, Papalexopoulos and Hesterberg, 1990], ARMA [Huang and Shih, 2003, Taylor, 2010], neural network [Hippert et al., 2001], fuzzy logic [Pandian et al., 2006], and support vector machine [Chen et al., 2004, Fan et al., 2009, Sapankevych and Sankar, 2009].

Table 6.1: F-measure for each method with 1 hour and 15 minutes aggregation, with 2 to 16 symbols, using Random Forest (RF), J48 decision tree, Naïve Bayes (NB), and Logistic Regression. + means that the encoding use a single lookup table for all houses.

	RF	J48	NB	Logistic	RF+	J48+	NB+	Logistic+
distinctmedian 1h 2s	0.44	0.42	0.52	0.41	0.33	0.31	0.32	0.33
distinctmedian 1h 4s	0.42	0.32	0.41	0.51	0.45	0.29	0.41	0.33
distinctmedian 1h 8s	0.45	0.56	0.70	0.43	0.42	0.47	0.51	0.35
distinctmedian 1h 16s	0.67	0.49	0.81	0.74	0.42	0.41	0.65	0.46
distinctmedian 15m 2s	0.44	0.48	0.46	0.44	0.34	0.33	0.30	0.36
distinctmedian 15m 4s	0.46	0.37	0.49	0.45	0.40	0.40	0.46	0.36
distinctmedian 15m 8s	0.42	0.33	0.52	0.46	0.44	0.29	0.54	0.49
distinctmedian 15m 16s	0.44	0.52	0.69	0.68	0.38	0.40	0.66	0.50
median 1h 2s	0.52	0.61	0.57	0.51	0.41	0.45	0.40	0.36
median 1h 4s	0.72	0.58	0.79	0.72	0.50	0.52	0.50	0.33
median 1h 8s	0.86	0.65	0.91	0.69	0.39	0.38	0.52	0.27
median 1h 16s	0.92	0.72	**0.95**	**0.95**	0.45	0.41	0.61	0.50
median 15m 2s	0.65	0.58	0.64	0.43	0.49	0.37	0.30	0.22
median 15m 4s	0.63	0.64	0.80	0.70	0.51	0.49	0.66	0.50
median 15m 8s	0.78	0.44	0.86	0.86	0.53	0.47	0.75	0.58
median 15m 16s	0.73	0.57	0.94	**0.95**	0.66	0.46	**0.82**	0.70
uniform 1h 2s	0.22	0.22	0.22	0.24	0.14	0.14	0.14	0.14
uniform 1h 4s	0.30	0.28	0.28	0.31	0.28	0.22	0.28	0.30
uniform 1h 8s	0.50	0.38	0.52	0.36	0.34	0.25	0.15	0.22
uniform 1h 16s	0.56	0.44	0.57	0.43	0.36	0.39	0.48	0.33
uniform 15m 2s	0.27	0.16	0.23	0.23	0.25	0.19	0.20	0.25
uniform 15m 4s	0.39	0.38	0.37	0.34	0.25	0.17	0.26	0.27
uniform 15m 8s	0.51	0.51	0.47	0.30	0.26	0.24	0.26	0.31
uniform 15m 16s	0.52	0.52	0.58	0.49	0.39	0.30	0.42	0.21
raw 1h	0.66	0.53	0.37	0.29	0.66	0.53	0.37	0.29
raw 15m	0.65	0.57	0.36	0.36	0.65	0.57	0.36	0.36
raw 1sec	**0.71**	0.52	0.24	-*	**0.71**	0.52	0.24	-*

*) this values is not computed due to Java heap space issues with the Logistic algorithm.

However, all are applied to load on distribution level. In this experiment, we use support vector machine for regression to forecast (real value) residential level consumption.

Experiment result can be seen in Figure 6.8 and 6.9. The performances are measured by *mean average error* (MAE). Although our symbolic forecasting suffers from its inability to express the real values (and use the center of their range instead), its performance is comparable to real value forecasting. In some cases, our symbolic forecasting outperform real value forecasting (as in forecasting for house 1, 4, and 6 using Naive Bayes, and house 1, 3, 6 using Random Forest). For *median*, and *distinctmedian*, their segmentation is based on values' frequency (more segment in range with more values). This helps them to give more detail on dense value ranges and be more coarse (leave out unnecessary detail) one less dense ranges. For *uniform*, although its segmentation is not based on value range frequencies, its segmentation plays a role in filtering unnecessary fluctuation which often happens in residential consumption.

Looking at the promising performance of our symbolic representation, it opens a possibility of customized segmentation with background knowledge. Consider a simple example of an expert who is interested on two segmentation: *low* and *high* consumption, where *low* means consumption below a certain threshold, and *high* means consumption above a certain threshold.

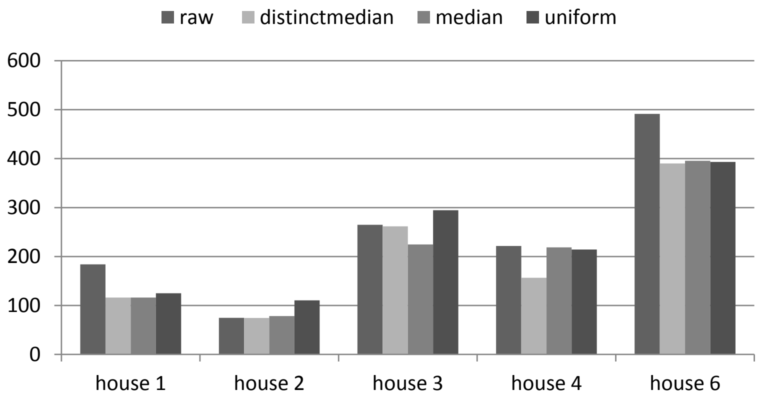


Figure 6.8: MAE of symbolic forecasting using Naive Bayes. Forecasting performance using raw value is shown for comparison. House 5 is skipped because there is not enough data.

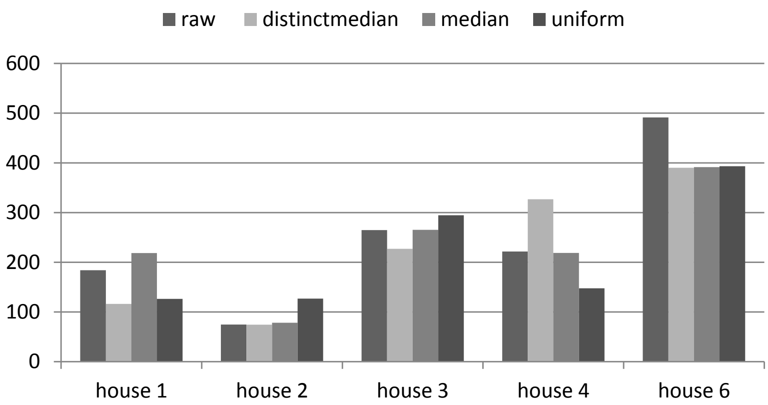


Figure 6.9: MAE of symbolic forecasting using Random Forest. Forecasting performance using raw value is shown for comparison. House 5 is skipped because there is not enough data.

This can be easily represented as a symbolic representation using alphabet of size 2.

6.4 Summary and Discussion

In this chapter, we have introduced three different methods to express smart meter data (hence, time series data) as symbols, namely: *median*, *distinctmedian*, and *uniform*. We aim to solve the data analytics problem in the smart grid due to very large data size generated by smart meter, which soon will exceed any consumption data generated so far, and its detail measurement which expose privacy risk to customers. Our symbolic representation reduce data numerosity and obscure smart meter detail measurements by representing them as symbols (where each symbol is a range of values).

We designed the horizontal segmentation with the goal of being flexible. Indeed, higher resolution symbols can easily be converted to lower resolution and lower resolution symbols can be compared to higher resolution ones. This allows to run most of the machine learning algorithms even if the symbolic time series have been encoded with different resolutions, or if the resolution changed in time.

The methods proposed in this chapter have been shown to perform well in experiments for data analytics on top of real data. Although their performances vary, depending on the algorithm used to perform the analytics tasks, in general they are comparable to the analytics performance of the real data. Experiment results show that the symbolic approach can be a promising direction to face the challenge of data numerosity and data privacy inherited by smart meter data.

Part II

Demand Response

7 Assessing Demand Response Baselines

Demand response (DR) baseline is a key factor in a successful incentive-based DR program since it influences the incentive allocation mechanism and customer participation. Previous studies have investigated baseline accuracy and bias for large, industrial and commercial customers. However, the analysis of baseline performance for residential customers has received less attention. This chapter analyzes DR baselines for residential customers. Our analysis goes beyond accuracy and bias by understanding the impact of baselines on all stakeholders' profit. Using our customer models, we successfully show how customer participation changes depending on the incentive actually received. We found that, in general, bias is more relevant than accuracy for determining which baseline provides the highest profit to stakeholders. Consequently, this result provides a valuable insight into designing effective DR incentive schemes.

This chapter has also appeared in IEEE Transactions on Smart Grid, vol.5, no.4 [see Wijaya et al., 2014d].

7.1 Introduction

Matching supply and demand is a key feature in the reliability of an electricity grid, since failure to ensure it could result in blackouts. Demand response (DR) can be seen as a demand side effort to match the available supply. This is essential, especially when there is nothing more that can be done from the supply side (particularly when energy sources are renewables).[1] This chapter focuses on incentive-based DR, where the incentive can be in the form of bill rebates, redeemable vouchers, discounts or any other monetary incentive.

There are two key factors to ensure the success of a DR program, namely (i) how to operate DR resources, and (ii) how to measure DR performance. The first factor depends on customers, the energy market, devices, and the utility company (or *company*). In this chapter, we focus on the second factor, and more specifically on the DR baseline (or simply *baseline*), which is an estimate of what customers would have consumed in the absence of a DR event. In an incentive-

[1] For more information about the benefit of DR, see [Strbac, 2008, U.S. Department of Energy, 2006].

based DR program, the baseline is important because it determines the incentives allocated to customers, and thus influences customers' decisions and participation.

Baseline Analysis Schnittger and Beare [2012] provided a brief overview of a DR baseline and its importance in a DR mechanism. Coughlin et al. [2008, 2009], KEMA, Inc. [2011], and EnerNOC [2011] analyzed the accuracy and bias of DR baselines, and suggested different adjustments to improve baseline accuracy. Our set of baselines is inspired by their work. However, we consider only the core baseline method, without adjustments. Their studies focused on accuracy and bias, and did not analyze how baselines affect stakeholders' profits. Mathieu et al. [2011] analyzed DR baseline error (hence, accuracy). However, instead of using a set of baseline methods, they considered only a regression-based baseline. They characterized baseline error using several parameters, and aimed to compute the error associated with each parameter. As in previous work, Mathieu et al. focused on baseline error/accuracy, and did not analyze further how a baseline affects the stakeholder profit.

Residential Demand Response Faruqui and George [2005] and Herter and Wayland [2007] quantified the effect of residential DR using critical peak pricing (CPP). Faruqui and Sergici [2010] found that peak demand could be reduced between 3% and 6% using time-of-use pricing (TOU) and between 13% and 20% using CPP. They compared customer's critical and normal weekday loads at the same temperature and found statistically significant average customer responses. In addition, a number of studies have proposed an automated response (aided by a software agent or energy management system) in reaction to the variability in energy prices (Section 2.2.1 and 2.2.3). Others have assumed the existence of small energy storage on customers' premises (Section 2.2.2). Overall, our literature review suggested that research in residential DR focused on dynamic pricing schemes. Investigations about incentive-based DR for residential customers has largely been ignored so far.

Overview of Contributions We summarize our contributions as follows. We study the impact of DR baselines applied to residential customers, whereas previous works focused on large, industrial and commercial customers [Coughlin et al., 2008, 2009, EnerNOC, 2011, KEMA, Inc., 2011]. While they concentrated on baseline accuracy and bias, we go beyond these by explaining how a baseline affects customer decision and participation in a DR event, and how it affects both customer and company profit. We evaluate existing methods as well as the new ones that we introduce here. We also develop three models of customer response during a DR event. We are able to model changes in customer participation as a response to the incentives they have actually received. We show that more positively biased baselines foster greater customer participation. Interestingly, while the idea of positively biased baselines does not work to the favor of utility companies, it does deliver the highest overall profits when profit sharing is low.

The rest of the chapter is organized as follows. In Section 7.2, we present different DR baseline methods. In Section 7.3, we provide the necessary definitions needed for our analysis, including customer and company profit. We present our analysis in Section 7.4 and 7.5. In Section 7.6, we conclude and outline the further implications of our work.

7.2 Demand Response Baselines

Consider a set of (residential) customers $\mathcal{C}$. We divide a day into a set of timeslots $\mathcal{T} = \{t_0, \dots t_{|\mathcal{T}|}\}$. We define the actual load of customer i on day (or date) d at timeslot $t \in \mathcal{T}$ as $\ell_i(d, t)$. In the presence of a DR event, we define the load that a customer would have consumed in the absence of a DR event as *true baseline*, i.e., the customer's intended consumption.[2] However, in practice, during a DR event the true baseline is unknown. Thus, to calculate a customer's reduction in demand (and her incentive for participation) during a DR event, the utility company needs to establish a DR baseline, or predicted baseline (or simply a *baseline* when the context is clear). *Predicted baseline* is the load that the utility company estimates a customer would have consumed in the absence of a DR event, i.e., the prediction of a customer's true baseline. We denote the predicted baseline of customer $i \in \mathcal{C}$ on day d at timeslot $t \in \mathcal{T}$ as $b_i(d,t)$, and her true baseline as $b_i^*(d,t)$.

Several methods have been proposed in the literature and used in practice to compute a predicted baseline load for a DR event: these include HighXofY, MidXofY, exponential moving average, and regression baselines [EnerNOC, 2011, KEMA, Inc., 2011]. For the completeness of our analysis, we also define a new baseline method: LowXofY. In addition, a DR baseline should be simple enough for all stakeholders to understand, calculate, and implement, including end-use customers [EnerNOC, 2011, KEMA, Inc., 2011]. Thus, even though more sophisticated machine learning methods could deliver higher prediction accuracy, we do not consider them in this study.[3]

We define *DR days*, or *target days*, as days when DR events occur, and others as *non-DR days*. Furthermore, we define two *day-types*: weekdays (Monday to Friday), and weekend (Saturday and Sunday). Let $D(Y,d)$ be a set of Y most recent non-DR days preceding the day d having the same day type as d. In addition, let $\ell_i(d) = \sum_{t\in\mathcal{T}} \ell_i(d,t)$ be the total load of customer i on day d.

7.2.1 HighXofY

HighXofY baseline considers Y non-DR days preceding the DR event. The baseline is the average load of the X highest consumption days within those Y days. More formally, for customer i, we define her HighXofY days preceding a DR event day d as $High(X,Y,d) \subseteq D(Y,d)$, where:

(i) $|High(X,Y,d)| = X$, and

(ii) $\ell_i(\hat{d}) \geq \ell_i(d')$ where $\hat{d} \in High(X,Y,d)$ and $d' \in D(Y,d) \setminus High(X,Y,d)$.

The HighXofY baseline of customer i for timeslot t on day d is

$$b_i(d,t) = \frac{1}{X} \sum_{d \in High(X,Y,d)} \ell_i(d,t). \tag{7.1}$$

Examples of HighXofY baseline are [KEMA, Inc., 2011]:

- PJM Economic: High4of5 for a weekday, and High2of3 for a weekend DR event.
- NYISO: High5of10 for a weekday, and High2of3 for a weekend DR event.
- CAISO: High10of10 for a weekday, and High4of4 for a weekend DR event.

[2] In the absence of a DR event, true baseline is equal to actual load.
[3] We refer interested readers to Chapter 4 and 5 for performance analysis of more sophisticated prediction methods.

7.2.2 LowXofY

We propose this new, yet relatively simple baseline method. Similar to the HighXofY baseline, the LowXofY baseline for day d is calculated using the X lowest consumption days of Y non-DR days (of the same day type) preceding d. More formally, for customer i, we define her LowXofY days preceding a DR event day d as $Low(X, Y, d) \subseteq D(Y, d)$, where:

(i) $|Low(X, Y, d)| = X$, and

(ii) $\ell_i(\hat{d}) \leq \ell_i(d')$ where $\hat{d} \in Low(X, Y, d)$ and $d' \in D(Y, d) \setminus Low(X, Y, d)\}$.

The LowXofY baseline of customer i for timeslot t on day d is

$$b_i(d, t) = \frac{1}{X} \sum_{d' \in Low(X,Y,d)} \ell_i(d', t). \tag{7.2}$$

From this baseline method, we use Low4of5, Low5of10, and Low10of20. Unlike previous baselines in HighXofY, for these three baselines we use the same configuration for X and Y for weekday and weekend DR events. As we will show later, this baseline method has higher accuracy, but has more negative bias than the others.

7.2.3 MidXofY

The MidXofY baseline for day d is calculated using X of Y non-DR days preceding d by dropping some of the lowest and highest consumption days, retaining only the X middle consumption days. Let $X, Y \in \mathbb{N}$, $X \leq Y$, and $(Y - X) \bmod 2 = 0$. In addition, let $Z = (Y - X)/2$. The MidXofY baseline is calculated using $D(Y, d)$ by dropping the Z-lowest and Z-highest consumption days. More formally, for customer i, we define her MidXofY days preceding a DR event d as $Mid(X, Y, d) = D(Y, d) \setminus (Low(Z, Y, d) \cup High(Z, Y, d))$. The MidXofY baseline of customer i for timeslot t on day d is

$$b_i(d, t) = \frac{1}{X} \sum_{d \in Mid(X,Y,d)} \ell_i(d, t). \tag{7.3}$$

For our analysis, we consider Mid4of6 (which has also been considered in KEMA, Inc. [2011]).

7.2.4 Exponential Moving Average

The exponential moving average baseline is a weighted average of a customer's historical load, where the weight decreases exponentially with time. This baseline is computed using historical load data from the beginning of the measurement day up to the day preceding the target day d.

Let $D(\infty, d) = \{d_1, \ldots, d_k\}$ be the set of all measured days preceding the target day d having the same day type as d. In addition, let $1 \leq \tau \leq k$ be a constant. We define $s_i(d_\tau, t)$ as the *initial* average load of customer i on timeslot t, i.e.,

$$s_i(d_\tau, t) = \frac{1}{\tau} \sum_{j=1}^{\tau} \ell_i(d_j, t) \tag{7.4}$$

The exponential moving average for $\tau < j \leq k$ is

$$s_i(d_j, t) = \big(\lambda \cdot s_i(d_{j-1}, t)\big) + \big((1 - \lambda) \cdot \ell_i(d_j, t)\big) \tag{7.5}$$

Table 7.1: Summary of the baseline methods.

Baseline	Short description
HighXofY	average of the highest X of Y days
LowXofY	average of the lowest X of Y days
MidXofY	average of the middle X of Y days
Exp. moving avg.	weighted average of customer's consumption
Regression	linear regression of customer's consumption

note: all historical data considering only non-DR days preceding the DR event

where $\lambda \in [0, 1]$. Then, we define the exponential moving average baseline for customer i on day d at timeslot t as:

$$b_i(d, t) = s_i(d_k, t). \tag{7.6}$$

The baseline for days earlier than $d_{\tau+1}$ is undefined.

For this baseline method, we consider the ISONE baseline [KEMA, Inc., 2011] where $\tau = 5$ and $\lambda = 0.9$. The ISONE baseline is undefined for a customer who joined the DR program for less than 5 days. Even though in practice the ISONE baseline is not applied to the weekend, in this chapter, we also compute the ISONE baseline for the weekend.

7.2.5 Regression

The baseline of day d is computed using linear regressions whose parameters are inferred on the basis of historical data taken from $D(Y, d)$. This method uses one linear regression predictor for each timeslot during the day, i.e., the baseline of customer i on day d at timeslot t is computed by:

$$b_i(d, t) = (\boldsymbol{\theta}_{i,t})^T \boldsymbol{x}_{i,t} + \epsilon_{i,t} \tag{7.7}$$

where $\boldsymbol{x}_{i,t}$ is the feature vector, $\boldsymbol{\theta}_{i,t}$ is the (vector of) regression coefficient, and $\epsilon_{i,t}$ is the error term. The feature vector is a vector of explanatory variables such as historical consumption, temperature, or sunrise/sunset time. Then, we estimated the regression coefficient and the error term using ridge regression, although other estimation methods can also be used.

Because our dataset, as we will explain later, does not contain temperature or other measurements which could potentially be explanatory variables, we use historical consumption at the same hour of the day as feature vectors. In order to capture the weekly trend, we set the length of the feature vector to 7 for weekday estimation, and to 5 for weekend days.[4] We consider two regression baselines:

- Reg1, where $D(Y, d)$ contains all historical data available,
- Reg2, where $Y = 150$.

Table 7.1 summarizes the baseline methods explained in this section.

[4] Note that from a specific weekday (or weekend day), to reach that same day one week ago, we need to go back at least 5 previous weekdays (or 2 previous weekend days).

7.3 Residential Demand Response

With large, industrial and commercial customers, DR is often contract-based. For example, customers agree to respond to a fixed number of DR events per year. However, DR among residential customers can be very dynamic and need not be contract-based. One of the common execution scenarios, which we also consider in this chapter, is as follows:

1. The company sends a DR signal to the customers.
2. Each customer decides whether she would like to respond to the signal or not.
3. Using customer's smart meter data, the company reads her actual load and calculates her incentive.

7.3.1 DR Signal

Below are two possible types of DR signal for residential customers:

1. A signal which communicates the DR event start/end times and the amount of kWh to be reduced.
2. A signal which communicates the DR event start/end times and lets the customer decide how much she is willing to reduce.

In order to understand how baselines and incentive allocation influence customers' decisions to reduce their consumption, we focus on DR signals of type 2.

7.3.2 DR Event

We define a DR event as a tuple $\delta = (\delta_{start}, \delta_{end})$ where δ_{start} is the start time and δ_{end} is the end time of the event, and we denote $d(\delta)$ as the day/date of the event. In addition, we define the total actual load, predicted baseline, and true baseline of customer i during event δ as:

$$l_i(\delta) = \sum_{t=\delta_{start}}^{\delta_{end}} l_i(d(\delta), t), \tag{7.8}$$

$$b_i(\delta) = \sum_{t=\delta_{start}}^{\delta_{end}} b_i(d(\delta), t), \tag{7.9}$$

$$b_i^*(\delta) = \sum_{t=\delta_{start}}^{\delta_{end}} b_i^*(d(\delta), t). \tag{7.10}$$

For our analysis (later in Section 7.5), we obtained customers' true baselines from the real-world dataset, and we model customers' responses during a DR event to generate the actual load. Figure 7.1 provides a simple illustration of the customer's actual load, predicted baseline, and true baseline when there is a DR event from 17:00 to 20:00.

Then, for an event δ, we define the aggregate actual load, predicted baseline, and true baseline over all customers as:

$$L(\delta) = \sum_{i \in \mathcal{C}} l_i(\delta). \tag{7.11}$$

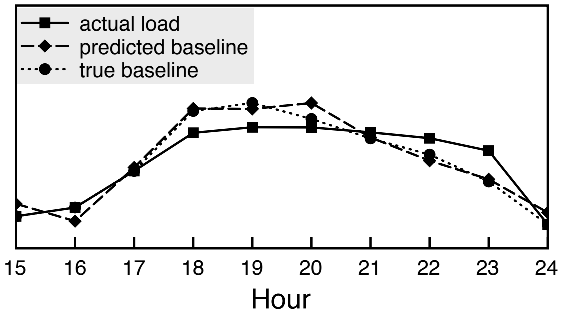


Figure 7.1: An illustration of the true baseline, predicted baseline, and actual load where a DR event occurs from 17:00 to 20:00 to curtail the evening peak.

$$B(\delta) = \sum_{i \in \mathcal{C}} b_i(\delta), \tag{7.12}$$

$$B^*(\delta) = \sum_{i \in \mathcal{C}} b_i^*(\delta), \tag{7.13}$$

In practice, while $B^*(\delta)$ is not known, $B(\delta)$ and $L(\delta)$ are known. Publishing this information does not violate customer privacy (since both of them describe information aggregated over all customers). In addition DR performance feedback can also be useful to foster customer participation [Faruqui et al., 2010]. Thus, we assume that the utility company publish $B(\delta)$ and $L(\delta)$.

7.3.3 Cost and Profit Functions

Cost We denote $c(L)$ as the total cost of meeting load demand L. We assume that c is monotonically increasing and strictly convex. An example of a real energy cost function that satisfies both the above assumptions is the quadratic cost function for thermal generators [Gonzalez Chapa and Vega Galaz, 2004, Mohsenian-Rad et al., 2010, Rao, 2006]:

$$c(L) = a_1 L^2 + a_2 L + a_3. \tag{7.14}$$

where a_1, a_2, and a_3 are constants.

Note that our assumption about the monotonically increasing cost function might not hold in cases of renewable energy generation, such as solar or wind power when there is an abundance of sunlight or wind. In this case, we consider a typical situation where there is a lack of supply, and more expensive generator or power sources need to be activated. This may involve advanced buying from the wholesale energy market at a few hours' notice.

A DR event typically happens when there is not enough supply to meet demand or when the spot market price is higher than the retail price. Therefore, the lower the customers' consumption, the greater the company's saving. We introduce the notion of a *company's true saving* as the differences between the cost of generating the customers' true baseline and the cost of generating the customers' actual load, i.e.,

$$c(B^*(\delta)) - c(L(\delta)). \tag{7.15}$$

However, in practice, what we can compute during a DR event is the predicted baseline, and not the true baseline. Under this condition, the company's saving is computed as the difference between the cost of generating the predicted baseline and the cost of generating the actual load. Hence, we define the notion of the *company's perceived saving*:

$$c(B(\delta)) - c(L(\delta)). \tag{7.16}$$

Customer's profit function Because of the customer's own efforts to reduce the load, and so as to give the customers further incentives, we define $\alpha \in [0, 1]$ as the proportion of the saving that the company would be willing to share with its customers. A customer only receives an incentive if she reduces her consumption in comparison to the predicted baseline. The incentive received is proportional to the aggregate load. We define the *received incentive* of customer i as:

$$rv_i(\delta) = \begin{cases} \alpha \cdot \big(\frac{b_i(\delta)}{B(\delta)} c(B(\delta)) - \frac{l_i(\delta)}{L(\delta)} c(L(\delta)) \big), & \text{if } l_i(\delta) < b_i(\delta) \\ 0, & \text{otherwise} \end{cases} \tag{7.17}$$

We can see from Eq. 7.17, that customer's received incentive depends on the predicted baseline established by the company. From the customer's perspective, if we know what she would have consumed (if there were no DR event), then we can also compute her *true incentive*, which is defined by replacing the predicted baseline in Eq. 7.17 with customer's true baseline:

$$tv_i(\delta) = \begin{cases} \alpha \cdot \big(\frac{b_i^*(\delta)}{B(\delta)} c(B(\delta)) - \frac{l_i(\delta)}{L(\delta)} c(L(\delta)) \big), & \text{if } l_i(\delta) < b_i^*(\delta) \\ 0, & \text{otherwise} \end{cases} \tag{7.18}$$

In this case, we assume that the customer is able to estimate her own intended consumption (true baseline). This could be made possible with the help of software agents for example, since the customer is the stakeholder with the most knowledge about the residence: the number and type of inhabitants, the number and type of appliances, and access to personal agendas to know when someone is at home or away. Because the company publishes only $B(\delta)$ and $L(\delta)$, $B^*(\delta)$ remains unknown to the customer. Thus, in Eq. 7.18 we use $B(\delta)$ as the approximation of $B^*(\delta)$. A customer's true incentive can also be thought of as the customer's received incentive when the predicted baseline established by the company perfectly estimates customer's true baseline.

In addition, for customer i, we define the difference between her received incentive and her true incentive for a DR event δ as her *additional profit*, i.e.,

$$rv_i(\delta) - tv_i(\delta). \tag{7.19}$$

Positive additional profit means that customer i receives more incentive than she deserves.

Company's profit function The company's profit can be calculated by subtracting the amount of incentives allocated to customers from the true or perceived savings. Note that in practice what it is possible to calculate is the company's perceived savings. However, the perceived savings depend on the chosen baseline method and does not reflect the company's true savings or losses after a DR event. Therefore, if possible, analyzing the company's true profit using its true savings is highly desirable. Below, we specify the company's true profit by its proportion to the cost of generating true baseline:

$$\frac{c(B^*(\delta)) - c(L(\delta)) - \sum_{i \in \mathcal{C}} (rv_i)}{c(B^*(\delta))}. \tag{7.20}$$

In our analysis, we compute the company's true profit by maintaining the customers' true baseline (obtained from a real-world dataset), and modeling customer responses during a DR event to obtain the actual load.

7.3.4 Customer Model

For our analysis, we propose three customer models.

Naïve This first model introduces a naïve customer model whose fixed parameter $\gamma \in [0, 1]$ determines by how much she reduces her intended load (true baseline). For customer i, for each DR event δ, we have:

$$l_i(\delta) = (1 - \gamma_i) \cdot b_i^*(\delta). \tag{7.21}$$

This parameter γ remains constant over time, regardless of the chosen baseline or the incentive given by the company.

Rational From Eq. 7.17 and 7.18, we can see that when the predicted baseline underestimates a customer's true consumption she receives less incentive than she deserves (and *vice versa*). In this model, a customer responds to a DR signal only if the predicted baseline established by the company does not underestimate her true consumption. More formally, for customer i, during a DR event δ:

$$l_i(\delta) = \begin{cases} (1 - \gamma_i) \cdot b_i^*(\delta), & \text{if } b_i(\delta) \geq b_i^*(\delta) \\ b_i^*(\delta), & \text{otherwise} \end{cases} \tag{7.22}$$

where $\gamma_i \in [0, 1]$ is the proportion of customer i's reduction compared to her true consumption.

Adaptive In this model, we introduce a customer who learns to make her decisions with regards to past experiences. The greater incentive the customer actually receives, in relation to the incentive she actually deserves, the more eager she will be to participate in the next event. That is, the customer's decision to reduce load evolves from one DR event to the next, influenced by the ratio between her received incentive (what she receives) and her true incentive (what she should have received).

Let δ_j be the jth DR event. From a predefined γ_i^0, this parameter evolves as the exponential moving average ratio between customer i's received incentive and her true incentive. Moreover, let

$$\begin{cases} \Delta^\omega = \frac{1}{\omega} \cdot \sum_{j=1}^{\omega} \frac{rv_i(\delta_j)}{tv_i(\delta_j)} \\ \Delta^j = \rho \cdot \Delta^{j-1} + (1 - \rho) \cdot \frac{rv_i(\delta_j)}{tv_i(\delta_j)}, & \text{for } j > \omega \end{cases} \tag{7.23}$$

where $\omega \in \mathbb{N}$ is the initial learning length parameter and $\rho \in [0, 1]$ is the decaying parameter to discount previous observations. Then, we define:

$$\begin{cases} \gamma_i^j = \gamma_i^0, & \text{for } j \leq \omega \\ \gamma_i^j = \Delta^{j-1} \cdot \gamma_i^{j-1}, & \text{for } j > \omega \end{cases} \tag{7.24}$$

We restrict the minimum value of γ_i^j to 0 and its maximum value to 1. In this model, customer i reduces γ_i^j of her true consumption during DR event δ_j, i.e.,

$$l_i(\delta_j) = (1 - \gamma_i^j) \cdot b_i^*(\delta_j). \tag{7.25}$$

The larger the ratio between customer's received incentive and her true incentive, the higher her γ for the next DR event.

7.4 Accuracy and Bias

7.4.1 Setup

For our analysis, we use the same Irish CER smart metering trial dataset as described in Section 3.5.1, which contains measurements of around 5,000 customers over 1.5 years. The customers

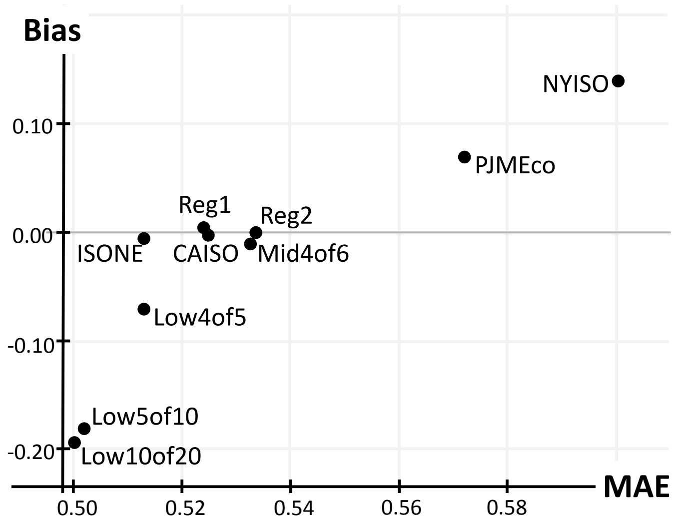


Figure 7.2: Mean Average Error (MAE) and bias of different baselines in kWh. Average hourly load over all customers is 0.97 kWh.

consist of residential houses and small and medium-sized enterprises. The measurements started in July 2009 and ended in December 2010. Since the trial was about dynamic pricing, similar to Chapter 3, we use only the data from the control group, composed of customers who are not affected by the different pricing schemes. More specifically, we choose residential customers that belong to the control group and have no missing values, resulting in the selection of 782 customers. In order to take into account the seasonal variation in customers' loads, we use a full year of measurement data, from January 1$^{\text{st}}$ to December 31$^{\text{st}}$ 2010.

We analyze the hourly accuracy and bias of each baseline. Let $\mathcal{C}$ be the set of our 782 customers, $\mathcal{D}$ be the set of all days in 2010, and $\mathcal{T}$ be the set of hourly timeslots in a day. We measure baseline accuracy in terms of Mean Absolute Error (MAE):

$$\frac{\sum_{i\in\mathcal{C}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}|b_i(d,t)-\ell_i(d,t)|}{|\mathcal{C}|\cdot|\mathcal{D}|\cdot|\mathcal{T}|}. \tag{7.26}$$

The lower the MAE, the higher the accuracy. And we define baseline bias as:

$$\frac{\sum_{i\in\mathcal{C}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}\big(b_i(d,t)-\ell_i(d,t)\big)}{|\mathcal{C}|\cdot|\mathcal{D}|\cdot|\mathcal{T}|}. \tag{7.27}$$

Baseline methods which have positive bias tend to overestimate customers' actual consumption (or true baseline when DR events occur), and *vice versa*.

7.4.2 Analysis

Figure 7.2 shows the accuracy and bias of each baseline in kWh. As expected, the HighXofY baselines whose $X < Y$, have a more positive bias than the others, whereas our LowXofY baselines have more negative bias than the others.

One interesting point here is that our LowXofY baselines are able to provide better accuracy than more sophisticated baselines such as ISONE (exponential moving average) and Reg1/Reg2

Table 7.2: Detailed characteristics of ISONE, Low4of5, Mid4of6, Reg2, and NYISO (ordered by MAE).

Baseline	Baseline family	MAE (kWh)	Bias (kWh)
ISONE	Exponential moving average	0.51362	-0.00259
Low4of5	LowXofY	0.51645	-0.07293
Mid4of6	MidXofY	0.53163	-0.01416
Reg2	Regression	0.53368	0.00893
NYISO	HighXofY	0.60611	0.14796

(regression).[5] The accuracy gained by our LowXofY baselines is driven by the days when a day with unusually high consumption is followed by days with normal (and thus much lower) consumption. In this case, LowXofY baselines will most likely exclude this unusually high consumption day from the next day's (or days') baseline computation(s). However, other baselines take this unusual day into account. For example, using CAISO, this unusual day will be carried through in the next 10 baseline computations (if the unusual day happens on a weekday, or 4 if it happens on the weekend). This is also the case with the exponential moving average and regression baselines, which both take into account this unusual day in their models.

7.5 Net Benefit Analysis

7.5.1 Setup

Baselines To analyze customer and company profit we focus on five representative baseline methods: ISONE, Low4of5, Mid4of6, Reg2, and NYISO. We chose them such that they use different baseline methods, and have different accuracy and bias profiles. Table 7.2 shows the details of their profiles.

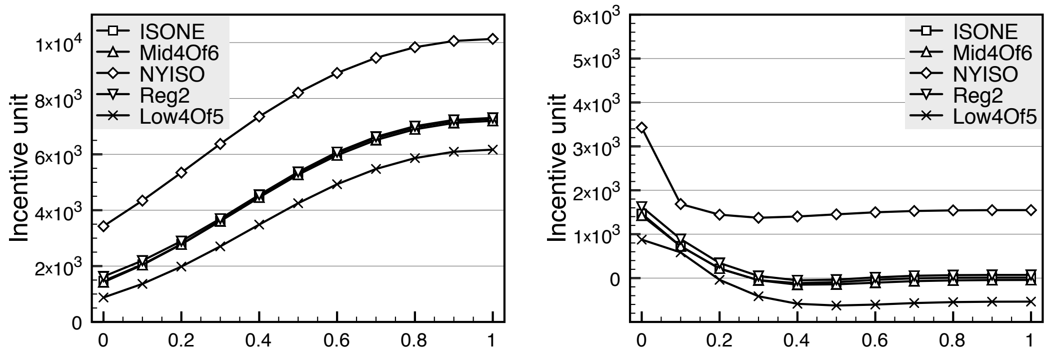
DR event We set the DR event to occur once a week on a random day. In total, we have 52 DR events for a year, between January 1st and December 31st 2010. The events happen evening during peak hours, starting at 17:00 and ending at 20:00. As described in Section 7.3.1, we use a type 2 DR signal, which allows us to analyze how the baseline and the incentive affect a customer's decision to reduce her load.

We use a simple cost function as described in Eq. 7.14 with $a_1 = 0.0001$, $a_2 = 0$, and $a_3 = 0$. Any other cost function could be used as long as it satisfies the assumption stated in Section 7.3.3. It will not affect our analysis (the end result might have different exact numbers, but would show the same trends). In addition, we assume that the imbalance between supply and demand occurs during the DR events. Hence, we focus our analysis exclusively on the time of the DR events.

7.5.2 Customers Profit

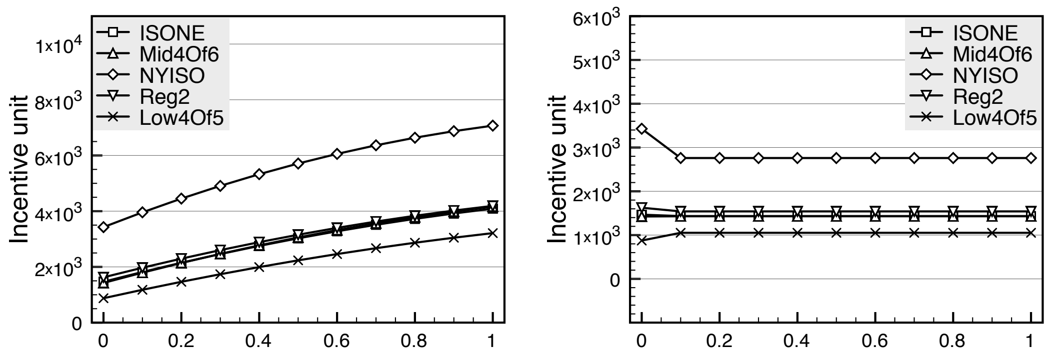
For each customer model, we analyze customers' received incentive and their additional profit when a particular baseline is used. Note that our cost function is not associated to a currency. Thus we define the unit of measurement of the customers' incentive as an *incentive unit.*

[5] However, one should not conclude that LowXofY is better than regression models in general. Adding some additional explanatory variables to Reg1/Reg2 could increase their predictive power and improve their accuracy.



(a) Received incentive (b) Additional profit

Figure 7.3: Received incentive and additional profit in the naive customer model with various γ (x-axis). Both are calculated as the sum of customers over all 52 DR events.



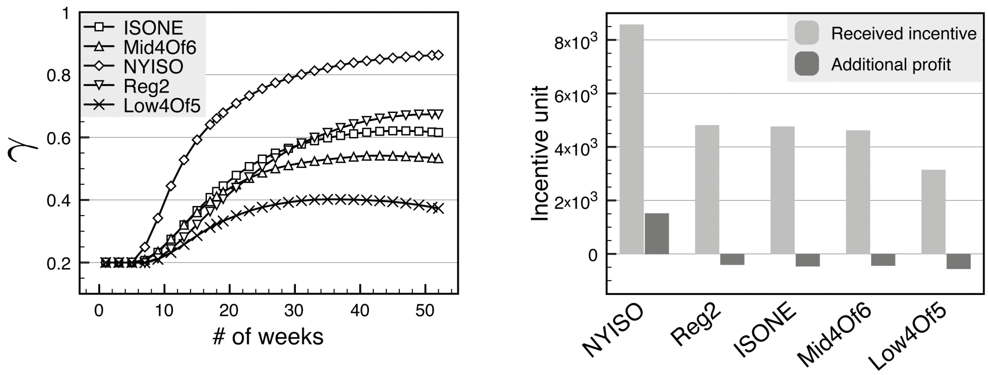
(a) Received incentive (b) Additional profit

Figure 7.4: Received incentive and additional profit in the rational customer model with various γ (x-axis). Both are calculated as the sum of all customers over all 52 DR events.

Received incentive in the naïve and rational customer models Figure 7.3a shows the received incentive in the naïve customer model, and Figure 7.4a shows the received incentive in the rational customer model. Both are shown as the sum of all customers over all 52 DR events with $\alpha = 0.1$ (defined in Section 7.3.3). We have the same trends, i.e., the larger the γ, the larger the received incentive. This is expected since γ represents the proportion of the intended consumption (true baseline) that customers reduce. A larger gamma means a lower actual load (Eq. 7.21 and 7.22), thus higher incentives (Eq. 7.17).

The figures show that irrespective of customers' γ, NYISO delivers them the highest incentives; Reg2, ISONE, and Mid4of6 come second; and Low4of5 delivers the lowest incentives. We highlight that this ranking is ordered by their bias (not accuracy). The more positive the bias, the higher we set the customers' predicted baseline loads. As a result, the customers received higher incentives (Eq. 7.17). When different α is used, our analysis does not change. Different α only introduces a constant shift to the current plots upward or downward (Eq. 7.17).

Additional profit of naïve and rational customer model Figure 7.3b shows the additional profit of the naïve customer model, and Figure 7.4b shows the additional profit of the rational



(a) Weekly average of γ (b) Received incentive and additional profit

Figure 7.5: The evolution of γ, received incentive, and additional profit of adaptive customer model with $\gamma^0 = 0.2$ and $\alpha = 0.1$ (defined in Section 7.3.3). Received incentives and additional profit shown are the sum over all customers and 52 DR events.

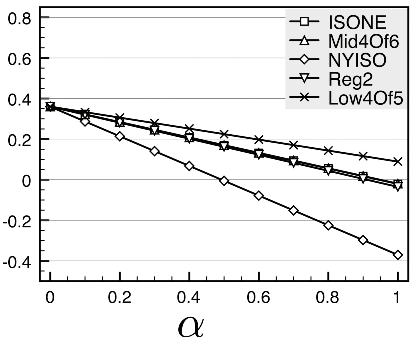
customer model. See Section 7.3.3 for the definition of additional profit. Similar to the received incentive case, baseline methods with more positive bias give higher additional profit to the customers. This can be understood since more positive bias baseline methods tend to overestimate the customers' true baseline, thus they give higher additional profit to the customers.

It is interesting to note how customers receive the highest additional profit when γ is low, especially when $\gamma = 0$ (except using the Low4of5 baseline on rational customers). Let $\mathcal{C}^+$ be the set of customers whose true consumption is overestimated and $\mathcal{C}^-$ be the set whose consumption is underestimated by a particular baseline for a particular DR event. When $\gamma = 0$, $\mathcal{C}^+$ has positive additional profit, whereas $\mathcal{C}^-$ has 0 additional profit. When $\gamma > 0$, $\mathcal{C}^+$ still has positive additional profit but $\mathcal{C}^-$ experiences negative additional profit, and the overall customers' additional profit decreases. In addition, when $\gamma = 0$ the true incentive of $\mathcal{C}^+$ is 0, whereas when $\gamma > 0$ the true incentive of $\mathcal{C}^+$ is > 0. This also causes the trend of additional profit going down when $\gamma > 0$. Additional profit can also be thought of as a "free lunch" for the customers.

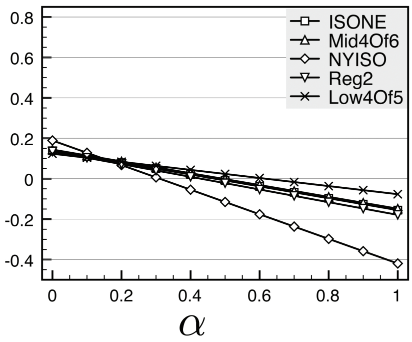
Customers with $\gamma = 0$ have a true incentive equal to 0, because they do not carry out any reduction in consumption. However, there are some customers whose loads are overestimated by the baseline methods, and others whose loads are not. While the non-overestimated customers do not receive any incentives, the overestimated customers do receive some incentives (free riders). This is why the total received incentive (and additional profit) over all customers is positive when $\gamma = 0$. For these customers, in order to receive bigger incentives, they need to increase their γ, i.e., they cannot be free riders any more. Even though their additional profits decrease, their received incentives increase (as the payoff for reducing their load).

Adaptive customers Figure 7.5 shows the evolution of customers' γ over time (weekly), their received incentives, and additional profits, given different baselines with initial gamma, $\gamma^0 = 0.2$. We recall that the evolution of a customer's γ depends on the ratio between her received incentive and her true incentive. The higher the ratio, the higher the customer's γ for the next DR event.

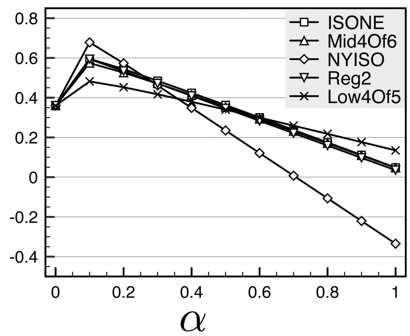
The primary motivation for the development of the NYISO baseline was to encourage customer participation. Using an adaptive customer model, we successfully realized this phenomenon. Figure 7.5a shows that using a positive bias baseline increases customers' γ more than using a negative bias baseline (see Table 7.2 for the bias of the baseline methods). This is due to the fact that more positively biased baselines provide a higher received incentives



(a) Naive customer model



(b) Rational customer model



(c) Adaptive customer model, $\gamma^0 = 0.2$

Figure 7.6: Company's profit under different customer models and various α.

compared to the true incentives.

The baseline bias and customer incentive trends seen in both the naïve and rational customer model can also be found in this customer models. A baseline with a more positive bias results in higher received incentive and additional profit (which encourages customer participation, as we mentioned earlier).

7.5.3 Company's Profit

Figure 7.6 shows the company's profit using different customer models. We discussed earlier that more positively biased baselines deliver higher customer profit. However, this is not the case with the company's profit. More negatively biased baselines provide higher company profit because they lower the amount of incentives allocated to the customers.

In the naïve customer model, and in the relatively high α of the rational and adaptive customer models, Low4of5 delivers the highest company profit compared to the other baselines. Moreover, we can see that the more negative the baseline's bias, the higher the company's profit. This is understandable since having more negatively biased baseline methods means that the company tends to set lower predicted baseline load, and hence distributes lower incentives to the customers. These lower total incentives lead to higher company profit. (see Eq. 7.20).

More interesting facts are shown in Figures 7.6b and 7.6c, which present some cases in which more positively biased baselines provide higher profit to the company. This is interesting because positively biased baselines (and, of course, larger α) are in line with customer preference. In general, negatively biased baselines, which tend to underestimate customers' true consumption, ought to deliver higher profit to the company. If this were always the case, then there would be a conflict of interest with the customers. Therefore, a case where more positively biased baselines deliver higher profit to the company is attractive because it opens the possibility of satisfying both stakeholders – the customers and the company.

However, determining the right α becomes crucial. Larger α results in a higher overall incentive allocated, thus, lower company profit. Figures 7.6b and 7.6c show that NYISO is the best baseline to use with the rational customer model with $\alpha < 0.15$ and with the adaptive customer model with $\alpha < 0.2$. In these two cases, for some small α, more positive bias baselines provide better profit for the company. This can be understood because a more positively biased baseline encourages more customer participation. Thus, it reduces the overall actual load, but does not give away too much in terms of incentives, which potentially increases the company's profit (see Eq. 7.20).

7.6 Summary and Discussion

In this chapter, we analyze the performance of DR baselines in the context of residential demand response, whereas previous works on baseline analysis focus only on large customers and commercial buildings. Furthermore, the baseline analyses performed to date were limited to "classic" analyses, i.e., baseline accuracy and bias. In this chapter, we go beyond these classic analyses by explaining the impact of DR baselines on the stakeholders' benefits, i.e., the profit of both the customers and the company. These are all essential elements for making residential DR a reality in the future.

As a supplement to the current baseline methods found in the literature, we propose a novel yet relatively simple baseline method: LowXofY. This method has a more negative bias than other baselines, but it is more accurate. We also successfully confirm the fact that positively biased baselines increase customer participation. While the motivation for using more positively biased baselines is indeed to encourage customer participation, little is known about whether overestimating customer consumption could benefit the company. We show that when the company shares a small portion of its profit with its customers (i.e., small α), it can actually increase its profit (and deliver higher profit compared to the other baselines) due to the increased customer participation. This opens up the possibility of a win-win solution for both the customers and the company. In addition, our result provides a valuable insight for the design of future incentive schemes for DR, i.e., even though the company can perfectly estimate a customer's baseline (which is useful to assess DR success rate), it might want to increase the baseline's bias a little to encourage her participation.

8 Matching Demand with Supply

Most of the works in Demand Response (DR) typically suggest reducing electricity generation cost by cutting the peak to average ratio (PAR), without necessarily reducing the total amount of the daily loads. However, most of these proposals rely on customer's willingness to act. This chapter proposes an approach to cut PAR explicitly from the supply side. The resulting cut loads are then distributed among customers by the means of a multiunit auction which is performed by an intelligent agent on behalf of the customer. Our approach is in line with the vision of the smart grid to have the demand side matched with the supply side and can also be applied to supply curves of any shape (e.g., due to high penetration of renewable energy sources).

This chapter has also been presented in the 5$^{\text{th}}$ International Conference on Communication Systems and Networks (COMSNETS) [see Wijaya et al., 2013c].

8.1 Introduction

Currently, electricity markets are designed so that the electricity supply has to fulfill the demand. When the demand increases rapidly, several problems occur, such as the possibility of power failures and high generation costs as expensive generators are turned on to fulfill demand for short peak periods. One vision for the smart grid has been to reverse this, and instead have demand match the available supply (see Section 1.1.2). This chapter proposes a method towards achieving this end.

The key to our method is the explicit cut of the peak to average ratio (PAR) of the electricity load generated. We introduce an approach which cuts PAR and then lets customers adapt by using an auction for redistributing the load. As others have proposed previously, smart home (automated, intelligent) agents could be used to represent customers in these auctions, making the entire process seamless from the customers' perspective [Oh and Thomas, 2008, Ramchurn et al., 2011, Vytelingum et al., 2011].

Both the challenges of obtaining low PAR and the advantages have been widely investigated [Cappers et al., 2010, Strbac, 2008, U.S. Department of Energy, 2006]. However, to the best of our knowledge, no other work proposes to explicitly cut PAR and let the demand side adapt to the supply. Many other approaches which have investigated cutting PAR have focused on having the demand side voluntarily adjust their consumption (given various incentives) in order to reduce peak load [Ganu et al., 2013, Li et al., 2011, Mohsenian-Rad et al., 2010, Wijaya et al., 2014e].

Using auctions in electricity markets has been both proposed in the literature and been used in practice, see for example [Bower and Bunn, 2001, Contreras et al., 2001]. In our work, however, we take a close look at the relationship between produces and customers and realize the vision of matching demand with supply while also considering demand satisfiability and minimum load guarantees, while also supporting the possibility of additional benefits for each side. The work most similar to that proposed here looked at the use of service curves [Le Boudec and Tomozei, 2011]. This work is the most similar to our work in the sense that it also imposes explicit restrictions on the customers' consumption curve. In [Le Boudec and Tomozei, 2011], however, the producers and customers has to agree in advance on the service curves contract, reducing the ability to adapt to changing supply conditions.

Overview of Contributions This chapter makes the following contributions:

- We propose an algorithm to explicitly cut PAR and prove its soundness and completeness.
- We introduce a (multiunit) auction for distributing load after the cut. We provide a minimum load guarantee for all customers, and show that the auction support truthful myopic bidding.
- Our simulations illustrate that our method benefits both customers and producers.

The rest of the chapter is organized as follows. We describe our model and other basic notions in Section II. In Section III we describe the algorithm for cutting PAR and prove that it is both sound and complete, while in Section IV we present our auction for distributing available load. Our experimental results are presented in Section V, which conclusions and further discussions in Section VI.

8.2 Preliminaries

8.2.1 Load Modeling

Load Let $\mathcal{N}$ be a set of customers and $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$ be a set of uniform time slots in a day. We denote the *electricity load* (or simply *load*) needed by customer $i \in \mathcal{N}$ in time slot $t \in \mathcal{T}$ as $L_i(t)$. Hence, the total load of customer i for a day is given by:

$$L_i = \sum_{t \in \mathcal{T}} L_i(t).$$

In addition, we define total load per time slot over all customers as:

$$L(t) = \sum_{i \in \mathcal{N}} L_i(t)$$

PAR Peak to average ratio (PAR) is a commonly used measurement to express how the peak compares to the average load:

$$PAR(L) = \frac{|\mathcal{T}|}{\sum_{t \in \mathcal{T}} L(t)} \max_{t \in \mathcal{T}} L(t) \tag{8.1}$$

8.2.2 Multiunit Auction

Before introducing our auction (in Section 8.4), in this section we briefly explain the basics of multiunit auctions.

Auction We use a *uniform price auction*, where each winning bidder pays the same price for each item/resource they win. The price paid is the price of the highest non-winning bidder. Given R resources and $\mathcal{N}$ as a set of agents or bidders, we denote a *bid* of an agent $i \in \mathcal{N}$ as a 2-tuple (r_i, v_i) where r_i is the number of resources desired by agent i and v_i is her *valuation* (the price i willing to pay for each resource she wins).

Winners The resources are won by the k highest bidders. Let $W \subseteq \mathcal{N}$ be the set of k agents who win the auction, then we require that:

(i) the winners are the highest bidders: $\forall i, j \ (v_i \geq v_j)$ where $i \in W$ and $j \in \mathcal{N} \setminus W$,

(ii) we allocate, tentatively, the maximum number of resources: if $W \subset \mathcal{N}$ then $\left(\sum_{i \in W} r_i\right) \geq R$, and

(iii) W is the smallest set of winners: $\forall W' \left(\sum_{i \in W'} r_i\right) < R$ where $W' \subset \mathcal{N}$.

Price The price paid by each winner $i \in W$ is the valuation of the highest non-winning bidder, i.e., $p^* = v_j$ where $\forall j'(v_j \geq v'_j)$ where $j, j' \in \mathcal{N} \setminus W$. Note that when $W = \mathcal{N}$ then there are two possibilities, either all bidders pay 0 or pay the reserve price (a price which is fixed by the auctioneer as the minimum price for a resource).

Resource distribution We sort the winning bidders by their valuation in ascending order. The resources are distributed to the winners starting from the highest bidders. Hence, there could be the case where the lowest winning bidder gets resources less than what she desires. In this case, she has an option to walk away (cancel her participation in the auction) or accept the resources offered.

Example 8.1 (Multiunit Auction). *A company would like to sell 6 resources. Then,*

- *bidder 1 would like to buy 2 resources at \$12,*
- *bidder 2 would like to buy 3 resources at \$10,*
- *bidder 3 would like to buy 3 resources at \$8, and*
- *bidder 4 would like to buy 1 resources at \$6.*
- *bidder 5 would like to buy 2 resources at \$5.*

Hence the winners of the auction are bidder 1, 2, and 3. Bidder 1, 2, and 3 received 2, 3, and 1 item respectively where each of them have to pay \$6 for an item. In this case, since the total demand of bidder 3 is not met, she can decide whether to take the 1 item offered and pay \$6, or withdraw from the auction (pay nothing and receive no item).

8.3 PAR-Cut

We cut $PAR(L)$ by a cut percentage, c, resulting in a new load vector L':

$$(1-c) \cdot PAR(L) = PAR(L'), \tag{8.2}$$

and

$$\sum_{t \in \mathcal{T}} L(t) = \sum_{t \in \mathcal{T}} L'(t). \tag{8.3}$$

When condition in Eq. 8.3 is met, using Eq. 8.1 we can rewrite Eq. 8.2 as:

$$(1-c) \cdot \max_{t \in \mathcal{T}} L(t) = \max_{t \in \mathcal{T}} L'(t). \tag{8.4}$$

In Algorithm 8.1 we provide a technique to explicitly cut the PAR of the original load generated by customer demand. The returned result is a load whose PAR has been cut. Cutting PAR while maintaining overall amount of load (as described in Section 8.3) means that there are some amount of load shifted from their original time slot. In this algorithm we aim to minimize the shift distance by first attempting to shift to a neighboring time slot.

There are several helper methods used in Algorithm 8.1:

- findPeak(L): returns the peak load of a load vector L.
- min(A, B): returns the smallest between two number A and B.
- shift(x, L, t_1, t_2): moves x amount of load from L_{t_1} to L_{t_2}.

Algorithm 8.1 receives input c as the cut percentage and L as the original load whose PAR is to be cut. We use Eq. 8.4 to define the new target peak p' (line 1). Then, for each time slot t_i, we verify whether the load at t_i exceeds p' (line 4). If this is the case, then we try to shift it (line 5-18). We first try to shift the load at t_i to its neighboring time slot by incrementing variable d for checking time slot t_{i+d} (line 10) and t_{i-d} (line 14). However, when we have investigated all time slots and there is still an amount of load to be shifted, then we conclude that it is not possible to do the cut, and the algorithm returns with failure (line 9).

Example 8.2 (PAR-cut)**.** *Let us assume that we have a set of time slots $\mathcal{T} = \{t_1, t_2, \ldots, t_{24}\}$, and load $L_{t_{19}} = 5kWh$, $L_{t_{18}} = L_{t_{20}} = 2kWh$ and $L_{t_i} = 1kWh$, where $1 \le i \le 17$, and $21 \le i \le 24$. In this case we have $PAR = 4$. Furthermore, let us assume that we want to cut the PAR by 40% cut percentage. Hence, our new peak $p' = 3$ (see line 1). First, Algorithm 8.1 detects that we have an excess load at time slot t_{19}, and the excess load $x = 2kWh$. Then, it proceeds with distributing the load to the neighboring time slot, t_{20} and t_{18}. This makes $L_{t_{18}} = L_{t_{19}} = L_{t_{20}} = 3kWh$. At the end of this step, the excess load $x = 0$, and the algorithm returns the newly modified load vector.*

For the proof of the soundness and the completeness of the algorithm, we refer to the input load to the algorithm as L, and the resulting load as L'.

Soundness. We need to show that if the algorithm returns L', then this is correct (having property shown in Eq. 8.3 and 8.4). Satisfying condition in Eq. 8.3 is straightforward since we never decrease or add new load. The only modification we apply is shifting the load from one time slot to another. Next, we always make sure that the maximum load in a time slot never exceeds p' (see line 4). This makes the condition in Eq. 8.4 hold. □

Algorithm 8.1: PAR-Cut

Input: cut percentage $0 < c \leq 1$, load vector L

1 $p' \leftarrow (1-c) \cdot \text{findPeak}(L)$
2 let $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$
3 **foreach** $t_i \in \mathcal{T}$ **do**
4   **if** $L[t_i] > p'$ **then**
5         $d \leftarrow 1$
6         $x \leftarrow L[t_i] - p'$ `/*excess load*/`
7   **while** $x > 0$ **do**
8   **if** $(i + d > |\mathcal{T}|) \wedge (i - d < 1)$ **then**
9   **return** *fail*
10   **if** $(i + d \leq |\mathcal{T}|) \wedge (L[t_{i+d}] < p')$ **then**
11                 $\textit{loadShifted} \leftarrow \min(x, p' - L[t_{i+d}])$
12                 $\text{shift}(\textit{loadShifted}, L, t_i, t_{i+d})$
13                 $x \leftarrow x - \textit{loadShifted}$
14   **if** $(x > 0) \wedge (i - d \geq 1) \wedge (L[t_{i-d}] < p')$ **then**
15                 $\textit{loadShifted} \leftarrow \min(x, p' - L[t_{i-d}])$
16                 $\text{shift}(\textit{loadShifted}, L[t_i], L[t_{i-d}])$
17                 $x \leftarrow x - \textit{loadShifted}$
18             $d \leftarrow d + 1$
19 **return** L

Completeness. We need to show that if the algorithm returns fail, then it is not possible to cut the load (as specified by the cut percentage c). The only condition where the algorithm returns fail is when it reaches line 9. This means that we still have an excess load, and we have no slot left with load less than p'. Hence, in order to satisfy the condition in Eq. 8.3, we have to put this excess load in some time slot. But this will make this time slot have load greater than p' which is not permitted by Eq. 8.4. □

8.4 Multiunit Auction for Load Distribution

In this section we explain how we run multiunit auction to distribute the cut load from Algorithm 8.1, or other load curves as long as they ensure the *minimal load guarantee* (Section 8.4.2). The auction is run either in the beginning of the day or a day before. It is held in several rounds until all loads are distributed. We use $\mathcal{N}$ as the set of customers and refer to agent i as the bidding agent representing customer $i \in \mathcal{N}$.

8.4.1 Initial Condition

Let L' denote the load vector on which we are going to run the auction, and let L denote the load vector that customers wish to obtain. For a time slot $t \in \mathcal{T}$, whenever $L'(t) \geq L(t)$, the demanded loads are distributed to the customers.

Let L_i^x be the *actual load vector* for customer $i \in \mathcal{N}$ (the actual loads delivered to customer i's residence) obtained in round x of the auction and let round 0 be the initial step before the auction begins. Then, we have customer i's actual load: $L_i^0(t) = L_i(t)$ for $L'(t) \geq L(t)$ and time slot $t \in \mathcal{T}$.

8.4.2 Minimal Load Guarantee

It could be a case that at time slot t there is a customer $i \in \mathcal{N}$ who does not get any load supply because all loads have been won by other customers. It means that at time t customer i does not have any electricity. This could happen when $L'(t) < L(t)$ for some $t \in \mathcal{T}$. That is, when we face a *short of supply* condition (because of the cut) where the load supply at time t is less than the actual customers' demand.

In order to prevent this condition from happening, we set m as the minimal amount of load guaranteed for each customer to have when we face a short of supply, i.e. for some time slot $t \in \mathcal{T}$ where $L'(t) < L(t)$, we require:

(i) $L_i^0(t) = \min(m, L_i(t))$ for all $i \in \mathcal{N}$,

(ii) $L'(t) \geq \sum_{i=0}^{|\mathcal{N}|} L_i^0(t)$.

Enforcing the minimal load guarantee above can also be seen as a mechanism to ensure the usage of non-shiftable appliances (such as fridge, lighting, multimedia or entertainment system). For customer i, the shiftable appliances (such as heating system, air conditioner, EV battery charger), however, need to be rescheduled by the smart home agent considering preferences of customer i, L_i, and L'_i.

8.4.3 Initial Pricing

To prevent the energy company getting paid 0 when we run the auction (this can happen when $W = \mathcal{N}$), we define the reserve price per Wh for each time slot $t \in \mathcal{T}$ as

$$p(t) = \frac{cost(L(t))}{L(t)}.$$

As in Section 7.3.3, we define $cost(L(t)) = a_1 L(t)^2 + a_2 L(t) + a_3$ to be increasing and convex, such that time slots having higher load is more expensive [see also Gonzalez Chapa and Vega Galaz, 2004, Mohsenian-Rad et al., 2010, Rao, 2006].

8.4.4 The Auction: Matching Demand Against Supply

We use the multiunit auction as described in Section 8.2.2 with some modifications:

1. An agent submits a set of bid $\{b_1, \ldots, b_{|\mathcal{T}|}\}$ all at once, where b_i is a bid for the load on time slot $t_i \in \mathcal{T}$.
2. Since it could be the case that not all loads are distributed to customers on one round of the auction, we run multiple rounds of the auction where the loads available for the next round are the loads left from the current round.
3. Bids (how much load an agent want and at what price) are placed simultaneously but separately for each time slot.
4. The winners are determined for each time slot.
5. The auction is terminated when there is no load left or there is no bid submitted.

Initial condition Let $L^x(t)$ be the loads left (after it has been distributed to the winners) after the x^{th} round of the auction. Thus, we denote the initial loads available before the auction begins as $L^0(t) = L'(t) - \sum_{i\in\mathcal{N}} L_i^0(t)$ for time slot $t \in \mathcal{T}$, where round 0 is the initial step before the auction begins and L' is the cut load vector.

Bid In each round x of the auction, agents place a bid on $L^{x-1}(t)$ using their own valuation to get some load they desired for each time slot $t \in \mathcal{T}$. However, if the load available in $L^{x-1}(t)$ is less than what an agent needs, then she will place a bid on $L^{x-1}(t')$ where t' is the closest time slot to t such that $L^{x-1}(t')$ is greater than what she needs. This is often called as *myopic best response* strategy, i.e. each agent is interested to maximize her own utility given the current condition. Note that it is possible that a sophisticated agent might be able to manipulate the system by reasoning and acting over several time steps and do a bit better. However, this involves sophisticated reasoning and could well be too complex for more simple bidders. And we will show that the optimal myopic strategy has attractive properties.

Winners determination The winners are determined separately for each time slot, i.e. we have a set of winners $W^x(t)$ for each time slot. Next, the resources available are updated to $L^x(t)$ by considering the load that has been distributed to the winners.

Demand satisfiability Let x^* be the final round of the auction. In addition, for a customer $i \in \mathcal{N}$ let $L_i'(t) = \sum_{x=0}^{x^*} L_i^x(t)$ be the sum of load she obtained through the auction for each time slot $t \in \mathcal{T}$. Then, we require for customer:

$$\left(\sum_{t\in\mathcal{T}} L_i'(t)\right) = \left(\sum_{t\in\mathcal{T}} L_i(t)\right).$$

That is, the total load obtained by a customer should be the same as her total electricity demand.

Truthful bidding is a myopic best response We prove that the auction supports truthful bidding, i.e. there is no incentive for agents to lie about their valuation.

Proof. Assume that there is still load that needs to be acquired by agent i on round x. Then, there are two cases:

1. $L^{x-1}(t)$ is equal or larger than what agent i needs. Then, agent i will bid on it. Assume that $v_i(t)$ is the agent i valuation for the load in time slot t. There are two cases:

   a) Agent i does not win. If agent i bid an amount $< v_i(t)$, it could not make her win. However, if she bid an amount $> v_i(t)$, she could win. But then she has to pay v' (the bid of the highest non-winning bidder), and it is also possible that $v' > v_i(t)$. In this case, agent i loss (because she has to pay higher than her valuation of the electricity at that time).

   b) Agent i win. Assume that the highest non-winning bid is v'. This is also the price that the winners have to pay. If agent i bid b', and $v' \leq b < v_i(t)$, then she still has to pay v' (she does not gain any additional benefit). However, if agent i bid $< v'$, then she lost. Hence, she does not get the load she want. And if agent i bid $> v_i(t)$, this also does not give any additional benefit because she still has to pay v'.

2. $L^{x-1}(t)$ is less than what agent i need. Then, agent i choose the closest time slot to t, that is t', such that $L^{x-1}(t')$ is equal or larger that what agent i need. Then, the same the argument as in the two cases above (1a and 1b), are applied here by replacing t with t'.

□

8.4.5 Customer's Load Shifting

Note that a bid for an amount of load at time slot $t \in \mathcal{T}$ placed by customer i does not necessarily comes from her actual need, from $L_i(t)$. Consider a simple example when customer i actually still needs to obtain her load for time slot $t' \in \mathcal{T}$ in the current round x, but $L^x(t') = 0$. Since there is no more load available in time slot t', then she has to place a bid for a different time slot $t'' \in \mathcal{T}$. When she wins the bid, load shifting for customer i occurs, i.e. she shifts an amount of electricity consumption from time t' to t''.

8.4.6 Customer's Utility

We define customer's utility after the auction by two terms:

(i) *cost*: we compute the cost paid by a customer at the end of an auction.

(ii) *shift percentage*: in order to measure a customer's inconvenience, we calculate the percentage of load that she has to shift. The shift percentage of customer i is:

$$\frac{\sum_{t \in \mathcal{T}} |L_i'(t) - L_i(t)|}{2 \cdot \sum_{t \in \mathcal{T}} L_i(t)}.$$

8.5 Experiments

In the experiments, we show the effects of the auction to customer's utilities for different type of customers. In addition, we investigate whether there is any additional benefit for the customer and/or the energy provider from running this auction.

8.5.1 Experiment Setup

In our experiments, we use an hourly time slot, i.e. $\mathcal{T} = \{1, \ldots, 24\}$, and total customers, $|\mathcal{N}| = 10000$. Then, we provide appliance usage setting to simulate customers' electricity demand. For the cost, any increasing and convex function can be used. In particular, we set $cost(L(t)) = \left((L(t) + q_1)/(q_2\sqrt{|\mathcal{N}|})\right)^2$ in order to include the energy company's marginal benefit, q_1, and to control the price's growth, q_2 (we set $q_1 = 100$ and $q_2 = 1000$).

For customer valuations (see Section 8.2.2), we use $\alpha \cdot p(t)$, where $p(t)$ is the reserve price at time slot $t \in \mathcal{T}$ (see Section 8.4.3). In order to set α among customers, we use two types of distributions

1. *US* distribution: for each customer, her α is drawn from a set $\{1.0, 1.3, 1.5, 1.6, 1.9\}$ with distribution 0.4, 0.2, 0.2, 0.1, and 0.1 respectively. This distribution is inspired by the US households wealth distribution in 2004 [1].

[1] http://www.faculty.fairfield.edu/faculty/hodgson/courses/so11/stratification /income&wealth.htm (accessed: 15 September 2012)

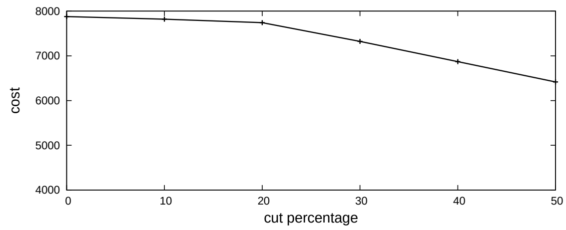


Figure 8.1: Overall system cost decrease as the cut percentage getting larger. Simulation using 10000 households loads.

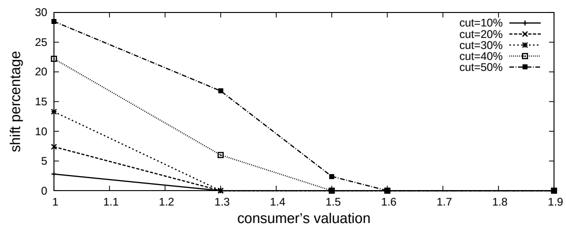


Figure 8.2: Customers' utility based on shift percentage measurement (defined in Section 8.4.6) grouped by their valuation (US distribution).

2. *Uniform* distribution: for each customer, her α is drawn uniformly from set $\{1.0, 1.1, 1.2 \ldots, 1.9\}$.

Each experiment is run 10 times and all results presented are with 95% confidence interval. However, the intervals almost cannot be seen because the deviations are very small.

8.5.2 System Cost

We present how the *PAR* cut affects the system cost. We show the result up to 50% cut percentage, since in our setting, it is not possible to cut the PAR with 60% or more. As expected, the more we cut the PAR, the lower the overall system cost is. Our experimental results in Figure 8.1 shows that cutting the PAR up to 20% does not give significant advantage. However, cutting the PAR with 50% cut percentage give almost 20% overall cost reduction. It looks very promising, but one have to be more careful and look on how customer's utility changes with increasing cut percentage (as we will show later in this section).

8.5.3 Customers' Utility

We measure customers' utility from two perspectives. First, we look at how much they need to shift their consumption. Second, we look at the total cost that they need to pay.

Shift percentage The results from both distributions (US and uniform) depicted in Figure 8.2 and 8.3 shows the same trends, i.e. customers with lower valuation are the ones who shift the most. This is expected since most of their demands which compete with higher valuation customers are not satisfied (because they lost in the auction). Hence they need to shift their consumption (bid for another time slot). For customers with valuation more than 1.7 we see no difference in their utility, and they do not shift at all. With this high valuation, we expect that they always win the auction and so see no difference in their supply.

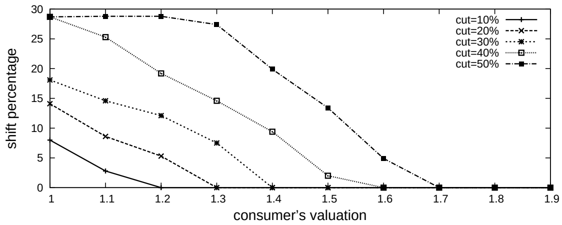


Figure 8.3: Customers' utility based on shift percentage measurement (defined in Section 8.4.6) grouped by their valuation (uniform distribution).

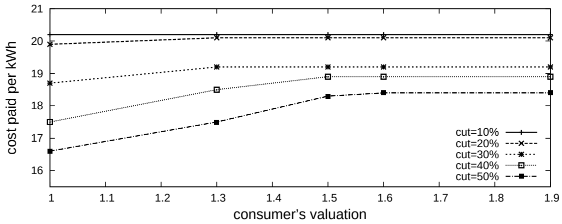


Figure 8.4: Customers' utility based on the total cost paid (defined in Section 8.4.6) grouped by their valuation (US distribution).

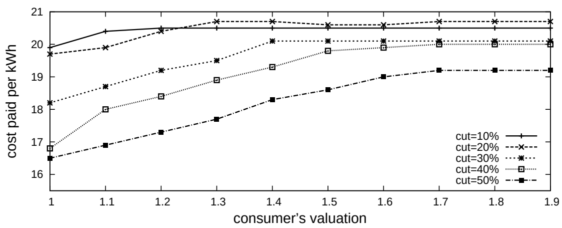


Figure 8.5: Customers' utility based on the total cost paid (defined in Section 8.4.6) grouped by their valuation (uniform distribution).

Cost paid Figure 8.2 and 8.3 both show the same trends that the higher the customer valuation, the higher the cost. This phenomenon is both caused by the reserve price and the auction itself. Most of the customers firstly will bid to satisfy their original consumption schedule, which happen most of the time in the peak time slot. Although we have already cut the peak, the price of this time slot is still the highest compared to the other (this is the time slot with the highest load). In addition, since most of the customers bid on this time slot, most likely the winners for this time slot will be the higher valuation customers, and the highest non-winning bid (which will determine the price they have to pay) will also be high.

As expected, low valuation customers pay lower costs. This can also be seen as the tradeoff (from the previous results) that they need to shift more than the customers with high valuation. The setting with 50% cut percentage offer lower cost paid compared to the others, but it also caused the low valuation customers to shift more (see Figure 8.2 and 8.3).

8.5.4 Customers' Additional Benefit

Compared to the current system, customers gain an additional benefit using the PAR cut and the auction. Customers can save up to almost 20% their electricity bill depending on the cut

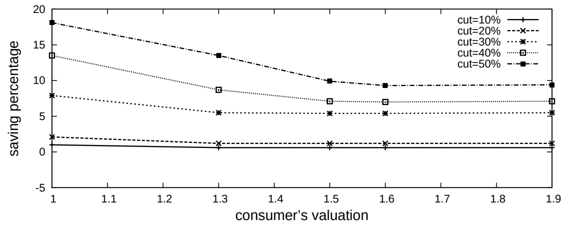


Figure 8.6: Customers' cost saving percentage using the auction and PAR cut compare to the current system. Valuation used: US distribution.

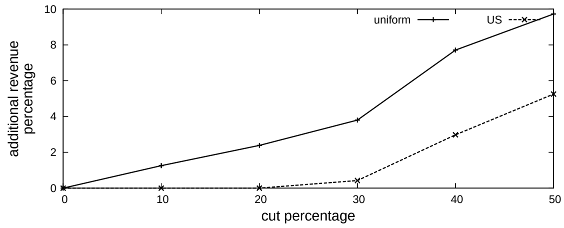


Figure 8.7: Company addition revenue from the auction.

percentage implemented (see Figure 8.6). This happens without having to reduce the amount of their electricity consumption. For customers with low valuation this saving can be seen as the tradeoff since they experience load shifting. However, what is more interesting is that this saving also applies to customers with high valuations (in this case, higher than 1.6) who experience no load shifting at all (see Figure 8.2).

8.5.5 Company's Additional Benefit

In this experiment, we calculated the total cost paid by customers and compared it to the energy company's cost model (the one that has to be paid by the customers, as described in Section 8.5.1). Figure 8.7 shows that the company experiences additional revenue by running the auction up to almost 10% depending on the PAR cut implemented. The larger PAR cut we have, the more peak loads are shifted. This will increase the price in the peak time slot because most likely the winners are customers with high valuations, and the highest non-winning bid will most likely also be high.

8.6 Summary and Discussion

In this chapter, we address the future vision of the smart grid to match the demand with the supply. Instead of relaying the PAR cut to the customer, we cut PAR explicitly while still maintaining the same load. Because the load available for each time slot are not necessarily the same as demanded by the customers, we use a multiunit auction to distribute the load. Additionally, since we do not make any assumption on the shape of L' and L, our multiunit auction can be used to have the demand side matched with the available supply curve of any shape as long as the minimal load guarantee (Section 8.4.2) is satisfied.

Our experiments show that the overall system cost is decreasing as we have larger PAR cut percentages. We also show that there is a tradeoff between the price per kWh paid by a customer with the load shift percentage she experienced. In addition, the customer with low valuation normally has bigger shift percentage and lower price per kWh than the customer with high valuation. Another advantage is shown for the customers that their total electricity cost is lower compare than their original electricity cost using the current setting (saving up to almost 20%). While for the low valuation customers this can be understood as the tradeoff for their load shifting, the saving also happens for high valuation customers who does not have any load shifting. The energy company also obtain an additional benefit from running the auction by acquiring up to almost 10% additional revenue. This result makes the proposed approach look promising.

9 Crowdsourcing Behavioral Incentives

In order to initiate pervasive adoption of Demand Response (DR) in the residential sector, the incentives need to be enticing, effective at promoting the desired energy consumption behavior, and able to maintain long-term consumer engagement. This chapter presents findings from a crowdsourcing experiment aimed at discovering innovative behavioral incentives for DR. The submission are then analyzed and classified according to Fogg's Behavior Model. We also show how the submission eventually can be combined into a more elaborate solution.

The preliminary version of this chapter can also be seen in [Wijaya et al., 2014c].

9.1 Introduction

Demand Response (DR) can be used for the purpose of demand regulation (e.g. to maintain voltage and frequency within safety limits) as well as for energy balance (e.g. to shift demand to off-peak periods, to curtail demand during emergency situations, or to offset fluctuations caused by less predictable energy sources such as wind or solar). Commercial and industrial energy customers have long became the main target of DR since they are able to contribute large reductions in demand through direct control of thermal loads (e.g. heating or refrigerators), higher predictability, lower user discomfort and relatively low installation costs. Although the residential sector makes up 20% of total energy demand and 60% of peak load demand, it still remains a relatively untapped DR resource.

A common DR approach aimed at changing energy consumption behavior is using dynamic pricing rates in which electricity tariffs are not flat, but fluctuate over time. Incentive-based programs are another variant of DR where customers receive benefits for participation in the program: these can be bill rebates or a discount rate when an amount of load reduction can be attributed to the customer during critical periods (see also Section 1.1.3). Although the two approaches are different in their design and operation, both rely on financial stimuli to affect consumers' energy consumption behavior. This chapter aims to find out various types of behavioral incentives that also can be used (and potentially more effective than the financial ones in some cases) to promote and maintain customer engangement in DR programs. For example, social interaction has been shown to have an impact on energy saving behavior [Brewer et al.,

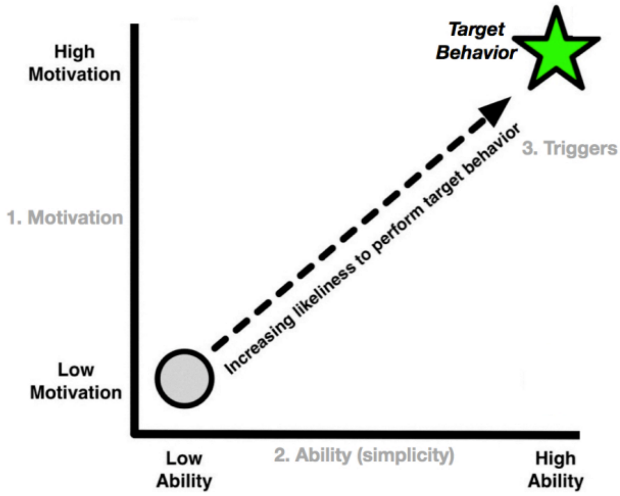


Figure 9.1: Fogg's Behavior Model (source: Fogg, 2009)

2011, Darby, 2006, Laskey, 2013, Mankoff et al., 2010, Petkov et al., 2011]. Additionally, practical experiments in energy conservation methods indeed found that informing a household about how energy usage compared to that of their neighbors' was much more effective in prompting them to conserve energy than informing them about the financial savings of lower energy usage [Cialdini and Schults, 2004].

Overview of Contributions We summarize the contributions of this chapter as follows.

- We conduct a crowdsourcing experiment aimed at discovering effective behavioral incentive mechanisms for promoting DR in residential customers (Section 9.3).
- We classify and analyze the submitted (behavioral incentives) ideas according to Fogg's Behavior Model (Section 9.4).
- Given that some ideas may be attractive to certain communities, but not others, new or hybrid solutions might need to be tailored specifically to suit the target community. In Section 9.5, we show how several submitted ideas can be combine into a more complete solution to sustain customer participation.

9.2 Behavioral Framework

To analyze and classify the behavioral incentives proposed by our crowd, we used Fogg's Behavior Model (FBM). According to this model, three factors must converge at the same time for a particular behavior to occur: motivation, ability, and trigger [Fogg, 2009]. The FBM asserts that a person must: (1) be sufficiently motivated to perform a target behavior; (2) have the ability to perform the behavior, and; (3) be triggered by some stimulus in order to actually perform the behavior (see Figure 9.1). FBM is introduced to analyze and design persuasive technologies.

9.2.1 Motivation

Motivation is a factor that can take various forms depending on the specific context of the target behavioral change. Fogg defined three different types of motivators:

Pleasure/pain This type of motivator is immediate and requires almost no thought, anticipation or planning. The response to this motivator is basically instantaneous, similarly to the response to hunger, sex or other activities related to biological self-preservation.

Hope/fear The second type of motivator requires a certain level of anticipation. The occurrence of a desired/undesired outcome is projected into the future, but makes the person experience the feeling of hope/fear in the present. For example, an energy conservation behavior can be motivated by the hope of reducing CO_2 emissions or by the fear of receiving an expensive electricity bill.

Social acceptance/rejection The third type of motivator is the social dimension. People are affected by social pressures which lead them to behave in ways that increase social acceptance and avoid social rejection. This motivator is highly present in virtual social networks, such as Facebook, where people actions (e.g. posting pictures, comments, etc.) are significantly driven by the desire for social acceptance.

9.2.2 Ability

Ability, or simplicity, is the factor of a persuasive mechanism that should reduce the effort needed to perform the target behavior. People are generally resistant to performing actions that require excessive effort, thus simplicity must be the guiding principle of any persuasive design. To be considered simple, a target behavior should be quick, cheap, require little physical or cognitive effort, should not violate social norms and conventions, and should easily become part of a person's normal routine. Of course, the absolute level of simplicity differs depending on the context of the persuasive mechanism to be put in place and users' profiles: some people have more time, or more money, or are more eager to perform non-routine actions than other people.

9.2.3 Triggers

Triggers are a very important factor of any persuasive mechanism. They can have various forms, use different communication channels, and must be recognized and associated with the target behavior. The FBM describes three types of triggers.

Spark This trigger is designed in combination with a motivator. Examples of sparks are text messages that highlight a fear, or videos that inspire hope.

Facilitator This trigger suits users who have high motivation but lack of ability, i.e. it is designed to improve the user's ability. The facilitator aims to trigger the behavior by making it easier to do. Examples of facilitators are software update messages that require only one click to install the update, or an address book uploading function on a social networking website that automatically builds an initial network of acquaintances on the user's behalf.

Signal This trigger is intended to work best when people have both the ability and the motivation to perform the target behavior: the signal is therefore a reminder. Examples of signals are automatic SMS or email messages that remind users to perform the target behavior, or traffic lights, which do not motivate the user but simply indicate when a behavior (i.e. crossing the intersection) is appropriate.

9.3 Crowdsourcing Experiment

9.3.1 Crowdsourcing

Crowdsourcing refers to the practice of requesting a large group of people (the crowd) to contribute to the accomplishment of a specific task [Brabham, 2013, Howe, 2006]. Crowdsourcing is usually offered by online services and platforms, and can be used for cloud labor, crowd creativity, crowdfunding, distributed knowledge and open innovation.[1]

Cloud labor consists on leveraging a distributed virtual labor pool on-demand to complete tasks of different complexities, from translating text to coding pieces of software. Amazon Mechanical Turk,[2] MobileWorks,[3] or Crowdflower[4] are well known platforms for this type of crowdsourcing. Crowd creativity is similar to cloud labor, but focuses on tapping pools of creative talent to develop art and media content, such as photography, video production, or graphic design. Crowdfunding seeks to raise financial contributions from a large number of stakeholders, sponsors, or donors to fund initiatives or enterprises. Kickstarter[5] is one of the most well-known platforms for starting a new business or enterprise. Distributed knowledge can be developed, aggregated, and shared through open questions and surveys, by using websites such as Epinions,[6] for example. Finally, open innovation refers to the generation, development, and implementation of ideas though brainstorming sessions. The best ideas are typically rewarded with some sort of prize or award. Websites that focus on this type of crowdsourcing are Ideascale[7] and Atizo[8].

Crowdsourcing is gaining acceptance in the innovation toolkits of many corporations and governmental agencies. A well-diversified crowd is composed of individuals with varied skills, experience, and perspectives, and can operate at a scale that often exceeds that of the biggest corporations. This means that the crowd can often solve problems more efficiently. For example, many corporations post predictive modeling and analytics competitions on Kaggle[9] so that statisticians and data scientists from all over the world can compete to produce the best models. This type of crowdsourcing approach is very effective for an exploration of the huge number of possible predictive models and techniques, given that it is usually impossible to know beforehand which ones will be most effective. Furthermore, crowdsourcing platforms are becoming cheap, powerful and easy to use, facilitating management of the task and interaction with the crowd members.

9.3.2 The Experiment

We performed a crowdsourcing experiment for open innovation and we asked the crowd for new ideas for effective incentive mechanisms for indirect DR. The challenge was posted on the Atizo website, the leading Swiss open innovation platform for the development of creative ideas and innovative concepts. On Atizo, a user (typically a corporation) can pose a challenge to be solved by others. A challenge comprises a description (short, catchy, and easy to understand), evaluation criteria, and reward.

[1] See also http://www.crowdsourcing.org/
[2] http://www.mturk.com
[3] www.mobileworks.com
[4] www.crowdflower.com
[5] www.kickstarter.com
[6] www.epinions.com
[7] www.ideascale.com
[8] www.atizo.com
[9] www.kaggle.com

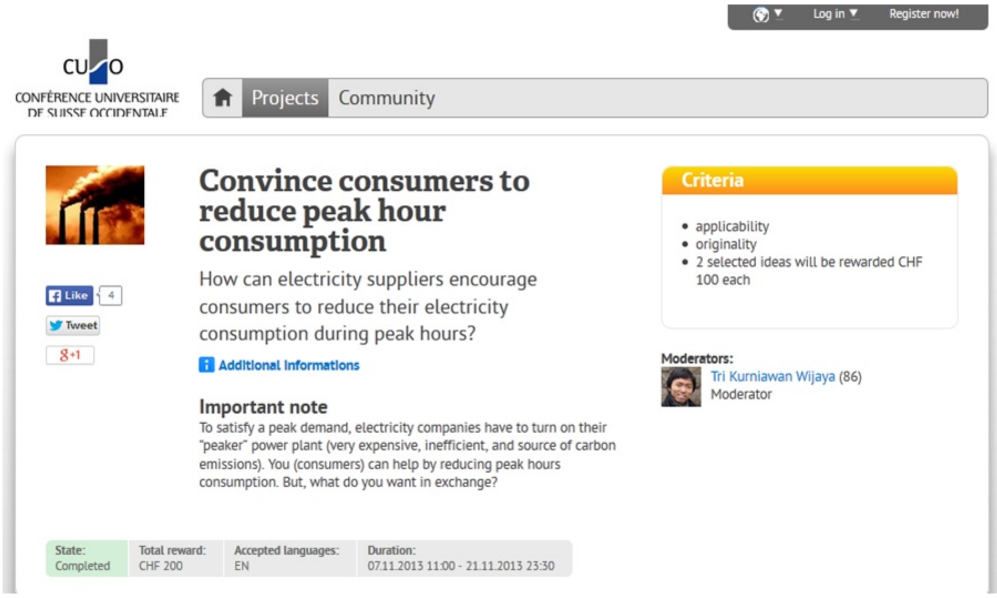


Figure 9.2: Home page of our crowdsourcing challenge

Our challenge was entitled "Convince consumers to reduce peak hour consumption. How can electricity suppliers encourage consumers to reduce their electricity consumption during peak hours?" To drive the originality and applicability of the ideas submitted, we used these factors as the principal evaluation criteria. The two best ideas were rewarded CHF 100 each. Figure 9.2 shows a screenshot of our challenge's home page.

9.4 Analysis of Results

9.4.1 Submission Statistics

The crowdsourcing experiment collected 55 ideas from 27 participants. The number of ideas submitted by each participant is shown in Figure 9.3. The histogram follows a power law distribution,[10] with the top three participants having provided 50% of all ideas. The campaign went almost unnoticed until the fourth day, when 27 submissions from 18 different participants were received. From that date onward, the crowdsourcing campaign steadily received an average of 3 new ideas per day (see Figure 9.4). The word cloud generated from the submissions (Figure 9.5) shows that they were indeed consistent with the challenge's main topic.[11] In the next sections, we provide an analysis of the submissions with respect to the three factors: motivation, simplicity, and trigger.

9.4.2 Motivation

The ideas submitted by the crowd participants allowed us to extract the following types of motivation that might drive energy consumers to reduce peak consumption:

[10] The best fit, showed as a solid line, is $f(x) = 8.7394x^{-0.7622}$

[11] See the complete list of the submissions at https://github.com/tritritri/behavioral-dr.

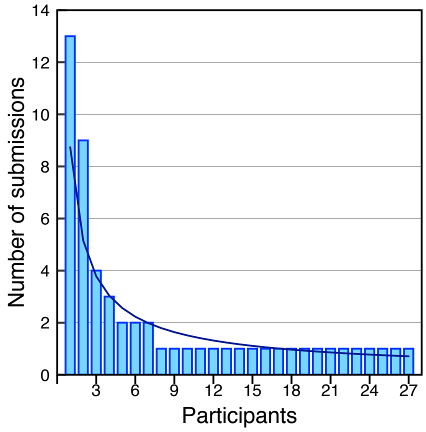


Figure 9.3: Number of submissions per participant

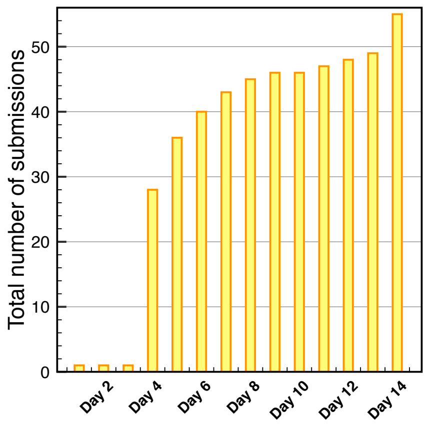


Figure 9.4: Number of submissions over time

Money (FBM category: Pleasure/pain). This motivation can take various forms, such as different tariffs between peak and off-peak consumption, discounts, bill rebates, redeemable points to use electricity at later time or to purchase energy efficient appliances.

Green awareness (FBM category: Hope/fear). People who care about the environment should be informed about the impact of their peak electricity consumption on CO_2 emissions and non-renewable resource usage.

Green cooperation (FBM category: Hope/fear). This motivation is similar to green awareness, but is enriched by a social component. Energy consumers are not aiming to reduce their personal carbon footprint, but rather they are members of a bigger community[12] aiming at more challenging goals, such as reducing peak consumption over an entire district and thus removing the need for an entire CO_2-intensive peak power plant. Consumers can

[12] Darby [2006] and Brewer et al. [2011] study the role of a community (village and dorm, respectively) in improving energy efficiency and conservation.

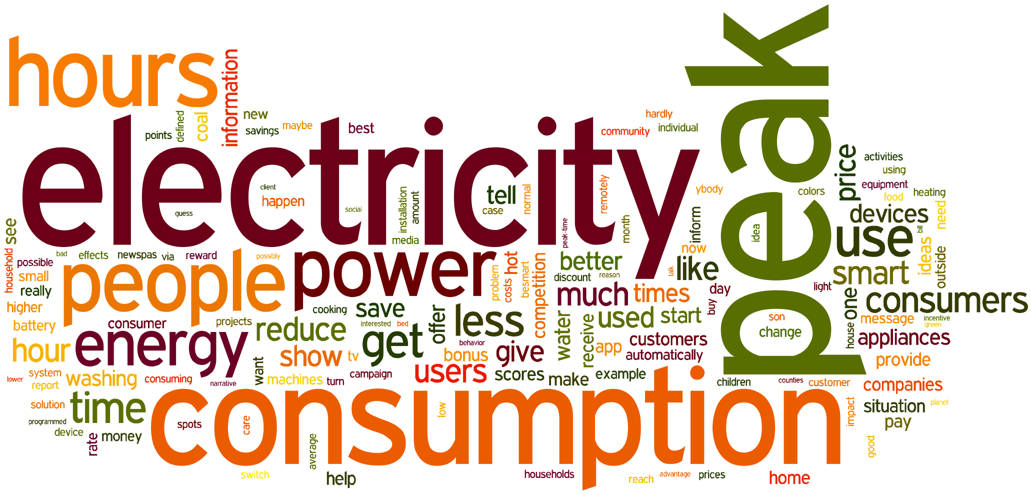


Figure 9.5: A word cloud generated from the submissions

also be motivated by the fact that the rewards for their collaborative effort are ploughed back into the community in the form of projects such as road maintenance or new parks.

Energy awareness (FBM category: Hope/fear). When consumers actually see how much they are consuming, a visual representation acts as the feedback to make them understand how and when they use energy. The understanding of their own energy behavior leads consumers to improve their peak energy consumption [Darby, 2008].

Social pressure (FBM category: Social acceptance/rejection). This motivation type refers to the pressure exerted by any type of social comparison or competition, including with oneself. For instance, the peak-shaving performance of energy consumers can be compared to that of other people in their neighborhood or shared within their social network using social media such as Facebook (see also [Mankoff et al., 2010]). This motivation can also take form of self-comparison, in which case people exert pressure on themselves.

A categorization of the motivational ideas submitted is shown in Figure 9.6. Money and energy awareness were identified as the strongest motivations. Protecting the environment, either as a personal challenge (green awareness) or as a collective effort (green cooperation), also represented a strong driver for behavioral change. Surprisingly, social pressure was not seen as a strong motivation, in contrast to the results of recent behavioral experiments. For example, Opower (www.opower.com), a company that partners utility providers in the promotion of energy efficiency, reported that leveraging competition between consumers was a strong driver in shaping energy consumption behavior; by showing them how their energy efficiency performance compared to that of their neighbors and other consumers with similar characteristics.

9.4.3 Simplicity

With regards to simplicity, participants in the crowdsourcing experiment highlighted the fact that energy consumers might need assistance to reduce the effort needed to reduce peak consumption. From their ideas, we extracted the following effort elements that consumers would like to have simplified (or minimized) in order to carry out the desired peak energy saving behavior:

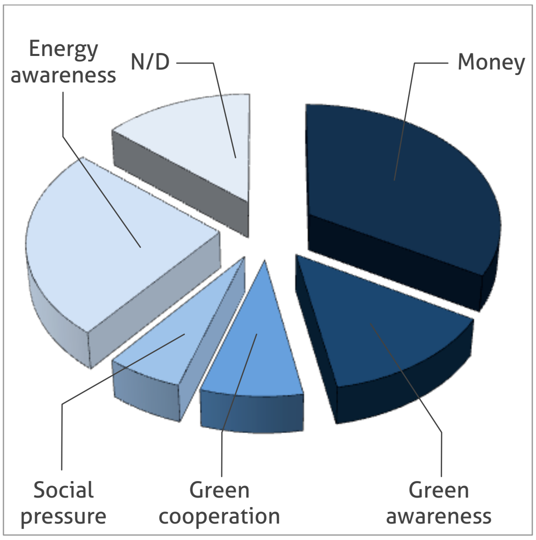


Figure 9.6: Motivation

Time The consumer often has both the motivation and knowledge to perform peak energy saving actions, but would like to spend the minimum amount of time on discovering when and which appropriate and specific actions should be carried out.

Physical effort All the ideas that require supportive devices such as smart appliances, programmable thermostats, electricity storage systems, home control systems and backup power generators, fall within a simplification of physical effort. All these devices reduce the physical effort required to carry out certain peak energy saving actions such as lowering the heating thermostat or rescheduling the use of appliances such as washing machines and dishwashers.

Cognitive effort In order to become savvy energy consumers, users must be educated. Any incentive that comes with informative content to help the user become an expert in energy management falls within this category.

A categorization of the ideas for simplicity submitted is shown in Figure 9.7. Although a considerable number of the submissions implicitly assumed that consumers were fully capable of carrying out the desired behavior, the majority of submissions recognized that helping consumers to reduce the cognitive and physical efforts necessary to carrying out that behavior was important for the incentive mechanism. Most residential energy consumers are neither tech savvy nor familiar with the their energy consumption, therefore educating them in their management of energy usage or teaching them to schedule appliances through easy-to-use control tools is a good way to increase their ability to perform the desired behavior.

9.4.4 Trigger

Finally, some of the ideas defined mechanisms aimed at triggering the desired consumer behavior. These triggers were classified into the following types:

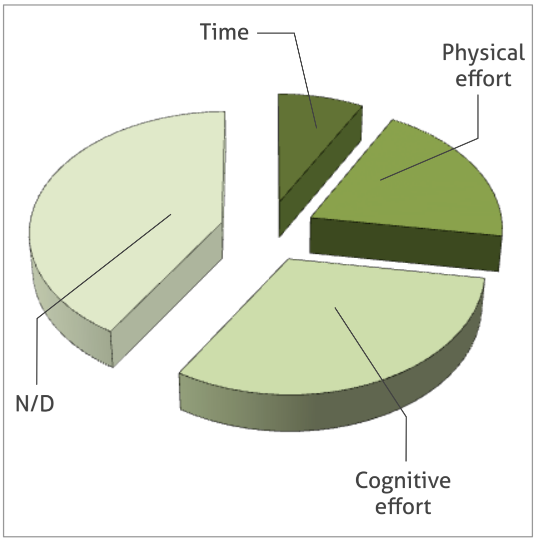


Figure 9.7: Simplicity

Media (FBM category: Spark). Many participants suggested the use of videos, documentaries, movies, text messages, radio announcements, and social media in order to instill motivations such as environmental and energy awareness.

Information (FBM category: Facilitator). This trigger aims to improve the consumer's ability to carry out the desired behavior by using appropriate graphical interpretations of peak energy consumption and possible actions to diminish it.

Alarm (FBM category: Signal). This category includes all types of reminders to carry out the desired behavior, and can be delivered via home displays, social media such as Twitter and Facebook, smartphone apps, and SMS messages.

A categorization of the trigger ideas submitted is shown in Figure 9.8. Almost half the contributions did not specify any sort of activation mechanism for triggering peak energy saving behavior. For the half that did consider triggers to be an important factor in any incentive mechanism, information dissemination campaigns using various media to encourage consumer motivation were considered the most effective. Furthermore, displaying information visually, using an appropriate graphical interpretation, was recognized as an important trigger to increase consumers' ability to carry out the desired behavior. Finally, only a small percentage of the submitted ideas specifically defined signaling triggers such as SMS alarms or reminders, sent through emails or social media such as Twitter and Facebook.

9.5 Crafting New Solutions

One of the advantages of carrying out an innovative crowdsourcing experiment is that aspects of the different ideas submitted can be combined to create new solutions. Given that some ideas may be attractive to certain communities, but not others, new or hybrid solutions can be tailored specifically to suit the target community. Furthermore, in order to move towards a more sustainable planet, behavioral change *per se* is not enough: new behavior should be sustainable

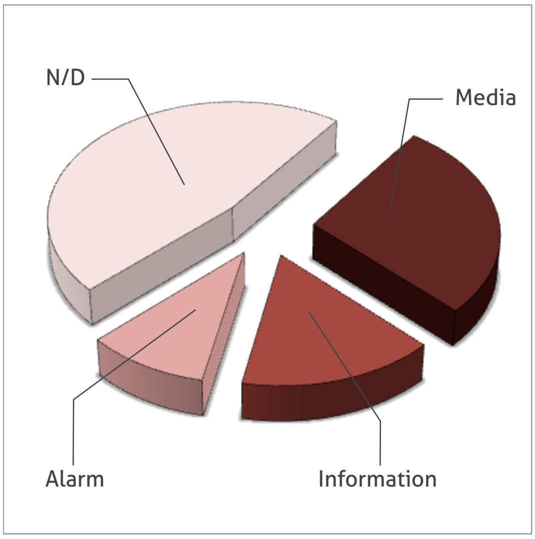


Figure 9.8: Triggers

in the long term. Below, we offer an example proposal that also aims to sustain change, where the overall objective is intentionally framed as a long-term endeavor in order to keep motivation high for the long run. We combine several crowdsourced ideas into one in order to define a behavioral incentive for DR which leverages both the participants' green awareness and their cooperative spirit as they collaborate to reach a long-term goal:

An advertizing campaign is run using TV ads and social media. People are asked to sign up for a challenging collaborative endeavor with the objective of reducing peak hour consumption at the scale of a community or even an entire city. A certain number of participants must be enrolled before the campaign can start, similarly to crowdfunding projects and companies. For a better mental image and intellectual grasp of the overall objective, it is set out as such a radical and definitive change in peak hour consumption that an entire CO_2-intensive peak power plant becomes unnecessary. The title of the campaign itself could be provocative, for example: "Let's Cut that Peak Power Plant!" *Participants should be able to join the campaign despite different levels of automation in their homes; from completely manual energy use management to programmable appliances, batteries, or even direct load control. The participants should receive easily understandable information about how to reduce consumption at peak time. Progress by the participants should be part of an appealing visual narrative (not exclusively charts or numbers) and could be deployed on several interfaces, from web-based applications to smartphone apps. Animated interpretations of progress, such as variably sized smoke clouds over the city, or animals and people in various states of health, could give a rapid emotional feedback to the participants in order to trigger further actions.*

9.6 Summary and Discussion

The challenge of influencing users to consume energy differently—to avoid peak energy consumption through demand response (DR), for example—must be met using elements of knowledge from power systems, information and communication technologies, economics, and behavioral science. Overcoming the multiple barriers to widespread adoption of DR in the residential sector will require advances in the design of behavioral incentive mechanisms that: (1) boost consumer motivation; (2) promote the desired energy behavior effectively, and; (3) maintain long-term consumer engagement. The analysis of the ideas collected in our crowdsourcing experiment showed that monetary incentives, albeit important, were not the only potential sources of consumer engagement. Consumers are indeed keen to protect the environment, either through individual acts or as part of a collaborative effort (see also [Gadenne et al., 2011]). Furthermore, a considerable percentage of the participants in our experiment clearly suggested (and validated the common belief) that consumers need mechanisms that reduce the cognitive and physical burdens of managing their energy consumption, such as through smart appliances, appropriate graphical interpretations of data, or decision support systems. Last but not least, whilst the Fogg Behavior Model (motivation, ability/simplicity, and trigger) can be used to model behavioral change, guaranteeing a more sustainable energy supply and environment also requires that the design of a DR program takes into account the long-term sustainability of the target behavior.

Part III

Conclusion

10 Conclusion

The electricity sector currently faces several important challenges, such as ever increasing demand (due to the growing world population and energy consumption per capita), and integration of renewable energy sources and electric vehicles (as an effort to reduce carbon emission and our dependency to fossil fuels). To this end, the emergence of smart grids is crucial in realizing a more sustainable energy system. Additionally, while it has always been the supply side that follows the demand, the demand side (customers) will soon have to be managed to match the available supply since the production of renewable energy sources are intermittent and unpredictable. This thesis focuses on understanding energy consumption behaviors of customers and mechanisms to influence or change those behaviors (i.e., through demand response) for the benefit of the whole system.

10.1 List of Key Findings

We summarize key findings of the thesis below.

- Customer segmentation results might vary among different contexts, e.g., segmentation in winter could be different from that in summer (see Figure 3.5 and 3.6a). Thus, to bring about an actual and up-to-date view of customer behavior, a versatile and context-based framework is needed.
- Different temporal aggregation of smart meter data (e.g., hourly, daily, monthly) matters depending on their application. Figure 3.6d shows that it matters for customer segmentation based on consumption variability or load shape (trends), but it does not matter for segmentation based on customer absolute consumption. It offers an important insight because, in general, less granular data is more efficient to store and process (more cost effective) and contains less private information (more privacy friendly) than highly granular data.
- Cluster evolution and individual behavior over time can be tracked using the cluster consistency index (Section 3.3).
- To maximize the competitive advantage from a customer segment (such as proposing a targetted/personalized services and products), one need to gain a deeper insight about the

segment. To this end, one can leverage the discriminative index (Section 3.4) to identify key customer characteristics that define a cluster.

- While floor area and the year a house was built is often discussed when estimating the house's energy consumption, we found that only floor area that matters (Figure 3.7). It could be explained by the fact that the older the house is the more likely it has been renovated by its owner, which makes the year it was built less relevant for predicting its building properties.
- The fact that the ownership of game consoles could be used as an indicator of a high-consumption household (Table 3.3 and 3.4) was rather surprising since game consoles consume only 100–200 watt (cf. dishwashers and tumble dryers that consume 1 kW or more). However, it can be explained by the fact that, first, the ownership of game consoles is highly correlated with family size. That is, the bigger the family is (e.g., with kids), the more likely it is to own game consoles. Second, the bigger the family is, the bigger its energy consumption. Both facts give rise to the correlation between the ownership of game consoles and household energy consumption.
- Smart meter data can be used to improve aggregate forecasting using the CBAF strategy (Section 4.5). We found that the improvement depends not only on the number of clusters but also on the size of the customer base.
- Instead of using bootstrapping or ensemble (which are computationally expensive) to develop prediction intervals of future demand, one could use GAM^2 algorithm (Section 5.2) to construct the intervals directly by estimating the time-varying conditional mean and variance, as long as the rescaled residual is approximately normal.
- The online learning algorithm in Section 5.3 can be used to adapt the prediction intervals to the ever changing trends in electricity demand.
- Turning smart meter data into symbolic representations (essentially aggregating them in a certain way) reduces its size and granularity, while still allowing analtyics to be performed on top of it (Section 6.2 and 6.3). It reduces storage cost and increases customer privacy (e.g., making Non-Intrusive Load Monitoring (NILM) less accurate), thus providing a trade-off between privacy protection and analytics accuracy.
- In incentive-based demand response (DR) programs, more positively biased DR baselines attract customer participation (Figure 7.5a, see Table 7.2 to compare the bias of various baselines).
- The bias of a DR baseline is more relevant to determine stakeholders' profit than its accuracy. In general, more positively biased baselines benefit customers (Figure 7.3, 7.4, and 7.5b), while more negatively biased baselines benefit utility companies (Figure 7.6).
- While more positively biased DR baselines typically do not benefit utility companies, Figure 7.6 shows that companies could benefit from them. They are, in fact, companies' best options when the profit sharing proportion is relatively small (i.e., ≤ 0.2 in the rational customer model and ≤ 0.3 in the adaptive customer model). In practice, intervention from policy makers might also be needed to protect customer interest.
- To reduce cost due to peak demand, we can explicitly cut the peak, shift them to the neighboring time slots (thus flattening the load curve), and distribute the supply using multiunit auction performed by smart home agents on behalf of customers (Section 8.3 and 8.4).

- The auction in Section 8.4 can be also used to distribute supply with curves of any shape. In practice, applying the auction to distribute supply generated by renewable energy sources would be of utmost interest in the near future.

- According to the Fogg's Behavioral Model, for a behavioral change to occur, a person must (1) be sufficiently motivated, (2) have the ability to perform the behavior, and (3) be triggered by some stimulus to perform the target behavior (Section 9.2). Thus, in addition to the more commonly applied monetary incentives, behavioral incentives in Section 9.4 can be used to maintain and sustain customer participation in DR programs.

10.2 What's Next?

Demand Analytics

In Chapter 3, we presented a versatile customer segmentation framework, which can also be used to understand customer consumption behavior, track behavioral changes, and find out key characteristics of a certain segment. When data about DR events is available, customer segments can also be defined based on responses to DR signals. Thus, characteristics relevant to the responses can also be identified. It would allow us to estimate the impact of future (planned) DR signals, which in turn enables us to develop DR design recommendation for delivering effective signals. The idea of using customer segmentation to help planning DR events has also been recently envisioned in [Chandan et al., 2014].

Chapter 4 discussed electricity demand forecasting for residential customers. Due to irregularities in household energy consumption, the smaller the number of households that we predict, the higher the focecasting error (see Figure 4.5, cf. [Tidemann et al., 2013, Figure 2-6]). It suggests that the potential and the outcome of DR through cooperatives (Section 2.2.3) might be easier to assess since energy consumption of a collection of households is more accurate to forecast than that of individual households. Additionally, in Chapter 5 we model the uncertainty of the future demand. The trade-off between the coverage accuracy and the interval width emerging from the adaptive construction algorithm could be further investigated. One could also study spatio-temporal dependencies and correlations between uncertainty of demand and renewable energy sources (thus, performing both, supply and demand analytics).

In Chapter 6, we introduced symbolic representations of smart meter data and dealt with contradicting goals of customers that aim to protect their privacy and utility companies that aim to perform more precise computations. Although we have explored some approaches for horizontal segmentation, there are still opportunities in developing more advanced strategies for both, horizontal and vertical segmentation, e.g., adaptive symbol table (to account for changes in consumption behavior), or utility-driven segmentation (optimized towards minimizing data size or maximizing computation accuracy, possibly with predefined properties or background knowledge from domain experts, see e.g., [Cao et al., 2014, Eberle et al., 2015]).

Demand Response

Chapter 7 studied how DR baselines affect stakeholders' (customers and utility companies) profit. Although the customer models do not take social interaction into account, social interaction has been shown to influence energy conservation awareness [Brewer et al., 2011, Laskey, 2013, Petkov et al., 2011] and results in a more environmentally-friendly behavior [Mankoff et al., 2010] (see also a result from our crowdsourcing experiment in Figure 9.6). Additionally, while the baseline methods described in Section 7.2 consider only non-DR days, a method that also take into account customer's consumption on DR days would be crucial in the future, especially when DR events are set to run almost every day. Moreover, baseline integrity (i.e., how far customers can manipulate a baseline estimation to their own advantages) is another important aspect of DR baselines that still needs to be assessed for a sustainable DR program.[1] Incentive distribution in the presence of encryption techniques to further secure the grid [Vasirani et al., 2014] is also another interesting avenue to explore.

Chapter 8 proposed an approach to match the demand with the available supply. It can be extended by considering cases where customers consider non-myopic strategies and reasoning over multiple steps at once. Thus, weak customers (that have low valuation) who reason over multiple steps might benefit from bidding on off-peak time slots (but possibly not so far from their preferences) for their needs in peak time slots. By bidding on off-peak time slots early, weak customers have better chances in avoiding competition with stronger customers (who also have to shift to those time slots because of even stronger customers). Additionally, the scheme can be applied not only to the cut supply (Section 8.3) but also to any other supply curves (especially those generated by renewable energy sources), as long as the minimal load guarantee (Section 8.4.2) can be fulfilled.

Following the crowdsourcing experiment in Chapter 9, in addition to monetary incentives, there are several other important behavioral incentives, e.g., energy awareness, social pressure, or green awareness. While studies about DR typically assume rational customers (as how it is normally assumed in economics), in reality customers often behave irrationally (which is also the reason *behavioral economics* emerges as a field of study). Therefore, in addition to financial incentives, DR designer should also consider behavioral incentives for motivating and sustaining customer engagement in DR programs. In the future, more studies on how psychological, social, and emotional factors affect customer responses during DR events might also be needed to complement the increasing number of computational studies in this area.

[1] Related work includes the work by Kota et al. [2012]. See also our discussion about it in Section 2.2.3.

Bibliography

Charilaos Akasiadis and Georgios Chalkiadakis. Agent cooperatives for effective power consumption shifting. In Marie desJardins and Michael L. Littman, editors, *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA*. AAAI Press, 2013. ISBN 978-1-57735-615-8.

Mohamed H. Albadi and Ehab F. El-Saadany. A summary of demand response in electricity markets. *Electric Power Systems Research*, 78(11):1989 – 1996, 2008. ISSN 0378-7796. doi: http://dx.doi.org/10.1016/j.epsr.2008.04.002.

Adrian Albert and Ram Rajagopal. Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, PP(99):1–12, 2013. ISSN 0885-8950. doi: 10.1109/TPWRS.2013.2266122.

Carlos Alzate and Mathieu Sinn. Improved Electricity Load Forecasting via Kernel Spectral Clustering of Smart Meters. In *The 13th IEEE International Conference on Data Mining (ICDM)*, pages 943–948, Dec 2013. doi: 10.1109/ICDM.2013.144.

Himanshu A. Amarawickrama and Lester C. Hunt. Electricity demand for Sri Lanka: A time series analysis. *Energy*, 33(5):724 – 739, 2008. ISSN 0360-5442.

Ning An, Weigang Zhao, Jianzhou Wang, Duo Shang, and Erdong Zhao. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy*, 49(0):279 – 288, 2013. ISSN 0360-5442.

Omid Ardakanian, Catherine Rosenberg, and S. Keshav. Distributed control of electric vehicle charging. In *Proceedings of the 4th international conference on Future energy systems*, e-Energy '13, pages 101–112, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2052-8. doi: 10.1145/2487166.2487178.

Amadou Ba, Mathieu Sinn, Yannig Goude, and Pascal Pompey. Adaptive learning of smoothing functions: Application to electricity load forecasting. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger, editors, *NIPS*, pages 2519–2527, 2012.

Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar Energy*, 83(10):1772 – 1783, 2009. ISSN 0038-092X.

Richard G. Baraniuk. Compressive sensing [lecture notes]. *Signal Processing Magazine, IEEE*, 24(4):118 –121, July 2007. ISSN 1053-5888.

Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. In *Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012)*, August 2012.

Christian Beckel, Leyna Sadamori, and Silvia Santini. Automatic socio-economic classification of households using electricity consumption data. In *Proceedings of the Fourth International Conference on Future Energy Systems*, e-Energy '13, pages 75–86, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2052-8. doi: 10.1145/2487166.2487175.

Beyond Zero Emissions. Smart Meters are a Smart Choice. https://bze.org.au/sites/default/files/images/Smart_Meters_-_BZE_Response.pdf, 2011.

Vincenzo Bianco, Oronzio Manca, and Sergio Nardini. Electricity consumption forecasting in Italy using linear regression models. *Energy*, 34(9):1413 – 1421, 2009. ISSN 0360-5442.

Nathaniel L. Bindoff, Peter A. Stott, Krishna Mirle AchutaRao, Myles R. Allen, Nathan P. Gillett, David Gutzler, Kabumbwe Hansingo, Gabriele C. Hegerl, Yongyun Hu, Suman Jain, Igor Mokhov, James Overland, Judith Perlwitz, Rachid Sebbari, and Xuebin Zhang. Detection and Attribution of Climate Change: from Global to Regional. In Thomas F. Stocker, Dahe Qin, Gian-Kasper Plattner, Melinda M.B. Tignor, Simon K. Allen, Judith Boschung, Alexander Nauels, Yu Xia, Vincent Bex, and Pauline M. Midgley, editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Laurent Bindschaedler, Murtuza Jadliwala, Igor Bilogrevic, Imad Aad, Philip Ginzboorg, Valtteri Niemi, and Jean-Pierre Hubaux. Track me if you can: On the effectiveness of context-based identifier changes in deployed mobile networks. In *Proceedings of the 19th Annual Network and Distributed System Security Symposium*, 2012.

Faycal Bouhafs, Michael Mackay, and Madjid Merabti. Links to the Future: Communication Requirements and Challenges in the Smart Grid. *IEEE Power and Energy Magazine*, 10(1): 24–32, Jan 2012. ISSN 1540-7977. doi: 10.1109/MPE.2011.943134.

John Bower and Derek Bunn. Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the england and wales electricity market. *Journal of Economic Dynamics and Control*, 25(34):561 – 592, 2001. ISSN 0165-1889.

Daren C. Brabham. *Crowdsourcing*. MIT Press, 2013.

Robert S. Brewer, George E. Lee, and Philip M. Johnson. The Kukui Cup: A Dorm Energy Competition Focused on Sustainable Behavior Change and Energy Literacy. In *The 44th Hawaii International Conference on System Sciences (HICSS)*, pages 1–10, 2011. doi: 10.1109/HICSS.2011.422.

Ciara Byrne. The smart grid data deluge. O'Reilly Radar, June 2011. URL http://radar.oreilly.com/2011/06/the-smart-grid-data-deluge.html.

José Ramón Cancelo, Antoni Espasa, and Rosmarie Grafe. Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting*, 24(4):588 – 602, 2008. ISSN 0169-2070.

Hông-Ân Cao, Christian Beckel, and Torsten Staake. Are domestic load profiles stable over time? an attempt to identify target households for demand side management campaigns. In *39th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, pages 4733–4738, Nov 2013. doi: 10.1109/IECON.2013.6699900.

Hông-Ân Cao, Tri Kurniawan Wijaya, and Karl Aberer. Estimating human interactions with electrical appliances for activity-based energy savings recommendations: Poster abstract. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, BuildSys '14, pages 206–207, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3144-9. doi: 10.1145/2674061.2675037.

Peter Cappers, Charles Goldman, and David Kathan. Demand response in u.s. electricity markets: Empirical evidence. *Energy*, 35(4), 2010.

Tommy Carpenter, Sahil Singla, Parsiad Azimzadeh, and S. Keshav. The impact of electricity pricing schemes on storage adoption in Ontario. In *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, e-Energy '12, pages 18:1–18:10, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1055-0. doi: 10.1145/2208828.2208846.

Vikas Chandan, Tanuja Ganu, Tri Kurniawan Wijaya, Marilena Minou, George Stamoulis, George Thanos, and Deva P. Seetharam. iDR: Consumer and grid friendly demand response system. In *Proceedings of the 5th International Conference on Future Energy Systems*, e-Energy '14, pages 183–194, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2819-7. doi: 10.1145/2602044.2602062.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

Mohamed Chaouch. Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. *IEEE Transactions on Smart Grid*, 5(1):411–419, Jan 2014. ISSN 1949-3053.

W. Charytoniuk, M.S. Chen, and P. Van Olinda. Nonparametric regression based short-term load forecasting. *IEEE Transactions on Power Systems*, 13(3):725 –730, aug 1998. ISSN 0885-8950.

Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen Lin. Load forecasting using Support Vector Machines: A study on EUNITE competition 2001. *IEEE Transactions on Power Systems*, 19 (4):1821–1830, 2004.

C. S. Chen, J. C. Hwang, and C. W. Huang. Application of load survey systems to proper tariff design. *IEEE Transactions on Power Systems*, 12(4):1746–1751, 1997. ISSN 0885-8950. doi: 10.1109/59.627886.

Gianfranco Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1):68 – 80, 2012. ISSN 0360-5442. doi: http://dx.doi.org/10.1016/j.energy.2011.12.031. 8th World Energy System Conference, {WESC} 2010.

Gianfranco Chicco, Roberto Napoli, Petru Postolache, Mircea Scutariu, and Cornel Toader. Customer characterization options for improving the tariff offer. *IEEE Transactions on Power Systems*, 18(1):381–387, 2003. ISSN 0885-8950. doi: 10.1109/TPWRS.2002.807085.

Robert Cialdini and Wesley Schults. Understanding and motivating energy conservation via social norms. Technical report, William and Flora Hewlett Foundation, 2004.

Javier Contreras, Oscar Candiles, José Ignacio De La Fuente, and Tomás Gómez. Auction design in day-ahead electricity markets. *IEEE Transactions on Power Systems*, 16(1):88 –96, Feb 2001. ISSN 0885-8950.

Katie Coughlin, Mary Ann Piette, Charles A. Goldman, and Sila Kiliccote. Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California. Technical Report LBNL-63728, Demand Response Research Center, CA, January 2008.

Katie Coughlin, Mary Ann Piette, Charles Goldman, and Sila Kiliccote. Statistical Analysis of Baseline Load Models for Non-Residential Buildings. *Energy and Buildings*, 41(4):374 – 381, 2009. ISSN 0378-7788.

Jack Danahy. The coming smart grid data surge. SmartGridNews, October 2009. URL http://www.smartgridnews.com/story/coming-smart-grid-data-surge/2009-10-05.

Sarah Darby. Social learning and public policy: Lessons from an energy-conscious village. *Energy Policy*, 34(17):2929 – 2940, 2006. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2005.04.013.

Sarah Darby. Energy feedback in buildings: improving the infrastructure for demand reduction. *Building Research & Information*, 36(5):499–508, 2008. doi: 10.1080/09613210802028428.

David L. Davies and Donald W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, 1979. ISSN 0162-8828. doi: 10.1109/TPAMI.1979.4766909.

Laurens James De Vries. *Securing the public interest in electricity generation markets. The myths of the invisible hand and the copper plate.* PhD thesis, Delft University of Technology, June 2004.

J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973. doi: 10.1080/01969727308546046.

Patty Durand. Market Segmentation: Defining A New Attitude For Utilities. http://smartgridcc.org/news-events/market-segmentation-defining-a-new-attitude-for-utilities, 2011.

Julien Eberle, Tri Kurniawan Wijaya, and Karl Aberer. Online Unsupervised State Recognition in Sensor Data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2015.

EnerNOC. The Demand Response Baseline. White Paper, 2011.

EU Commission Task Force for Smart Grids. Expert Group 1: Functionalities of smart grids and smart meters, 2010.

Jianqing Fan and Qiwei Yao. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645 – 660, 1998.

Shu Fan and Rob J. Hyndman. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1):134–141, Feb 2012. ISSN 0885-8950.

Shu Fan, Kittipong Methaprayoon, and Wei-Jen Lee. Multiregion load forecasting for system with large geographical area. *IEEE Transactions on Industry Applications*, 45(4):1452 –1459, July-aug. 2009. ISSN 0093-9994.

Ahmad Faruqui and Stephen George. Quantifying customer response to dynamic pricing. *The Electricity Journal*, 18(4):53 – 63, 2005. ISSN 1040-6190. doi: http://dx.doi.org/10.1016/j.tej.2005.04.005.

Ahmad Faruqui and Sanem Sergici. Household response to dynamic pricing of electricity: A survey of 15 experiments. *Journal of Regulatory Economics*, 38(2):193–225, 2010. ISSN 0922-680X. doi: 10.1007/s11149-010-9127-y.

Ahmad Faruqui, Sanem Sergici, and Ahmed Sharif. The impact of informational feedback on energy consumption – A survey of the experimental evidence. *Energy*, 35(4):1598 – 1608, 2010. ISSN 0360-5442. doi: http://dx.doi.org/10.1016/j.energy.2009.07.042.

Vera Figueiredo, Fátima Rodrigues, Zita Vale, and Joaquim Borges Gouveia. An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, 20(2):596–602, 2005. ISSN 0885-8950. doi: 10.1109/TPWRS.2005.846234.

Christoph Flath, David Nicolay, Tobias Conte, Clemens Dinther, and Lilia Filipova-Neumann. Cluster analysis of smart metering data. *Business & Information Systems Engineering*, 4(1): 31–39, 2012. doi: 10.1007/s12599-011-0201-5.

B. J. Fogg. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology*, pages 0:1–40:7, 2009.

Danielle Fong. Energy storage under pressure. TED, 2014. URL `http://tedxcern.web.cern.ch/video/2014/energy-storage-under-pressure-0`. Video file.

Erich Fuchs, Thiemo Gruber, Jiri Nitschke, and Bernhard Sick. Online Segmentation of Time Series Based on Polynomial Least-Squares Approximations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2232–2245, 2010.

Francesco Fusco, Michael Wurst, and Ji Won Yoon. Mining residential household information from low-resolution smart meter data. In *21st International Conference on Pattern Recognition (ICPR)*, pages 3545–3548, 2012.

David Gadenne, Bishnu Sharma, Don Kerr, and Tim Smith. The influence of consumers' environmental beliefs and attitudes on energy saving behaviours. *Energy Policy*, 39(12):7684 – 7694, 2011. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2011.09.002. Clean Cooking Fuels and Technologies in Developing Economies.

Tanuja Ganu, Deva P. Seetharam, Vijay Arya, Jagabondhu Hazra, Deeksha Sinha, Rajesh Kunnath, Liyanage Chandratilake De Silva, Saiful A. Husain, and Shivkumar Kalyanaraman. nPlug: An autonomous peak load controller. *IEEE Journal on Selected Areas in Communications*, 31(7):1205–1218, July 2013. ISSN 0733-8716. doi: 10.1109/JSAC.2013.130705.

Clark W Gellings. *The Smart Grid: Enabling energy efficiency and demand response*. The Fairmont Press, Inc., 2009.

M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, and M.S. Fadali. Smart meter based short-term load forecasting for residential customers. In *North American Power Symposium (NAPS)*, pages 1–5, Aug 2011.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, 2014. doi: 10.1146/annurev-statistics-062713-085831.

M. A. Gonzalez Chapa and J. R. Vega Galaz. An economic dispatch algorithm for cogeneration systems. In *IEEE Power Engineering Society General Meeting*, pages 989–994 Vol.1, 2004. doi: 10.1109/PES.2004.1372985.

Jan J. J. Groen, Richard Paap, and Francesco Ravazzolo. Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31(1):29–44, 2013. doi: 10.1080/07350015.2012.727718.

Christophe Guille and George Gross. A conceptual framework for the vehicle-to-grid (V2G) implementation. *Energy Policy*, 37(11):4379 – 4390, 2009. ISSN 0301-4215. doi: 10.1016/j.enpol.2009.05.053.

Tian Guo, Zhixian Yan, and Karl Aberer. An adaptive approach for online segmentation of multi-dimensional mobile data. In *Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, MobiDE '12, pages 7–14, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1442-8.

Martin T. Hagan and Suzanne M. Behr. The time series approach to short term load forecasting. *IEEE Transactions on Power Systems*, 2(3):785–791, Aug 1987. ISSN 0885-8950. doi: 10.1109/TPWRS.1987.4335210.

Mark Hall. *Correlation-based feature selection for machine learning.* PhD thesis, The University of Waikato, 1999.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009. ISSN 1931-0145.

Simon Harrison. Smart Metering Projects Map. http://tinyurl.com/smartmetermap, 2013.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*, volume 2. Springer, 2009.

Karen Herter, Patrick McAuliffe, and Arthur Rosenfeld. An exploratory analysis of California residential customer response to critical peak pricing of electricity. *Energy*, 32(1):25 – 34, 2007. ISSN 0360-5442. doi: 10.1016/j.energy.2006.01.014.

Andrea Hess, Francesco Malandrino, Moritz Bastian Reinhardt, Claudio Casetti, Karin Anna Hummel, and Jose M. Barceló-Ordinas. Optimal deployment of charging stations for electric vehicular networks. In *Proceedings of the first workshop on Urban networking*, UrbaNe '12, pages 1–6, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1781-8. doi: 10.1145/2413236.2413238.

Henrique Steinherz Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55, 2001.

Eric Hirst and Brendan Kirby. *Retail-load participation in competitive wholesale electricity markets.* Edison Electric Institute Washington, DC, USA, 2001.

Tao Hong. *Short-Term Electric Load Forecasting.* PhD thesis, North Carolina State University, September 2010.

Tao Hong, Pierre Pinson, and Shu Fan. Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2):357 – 363, 2014a. ISSN 0169-2070.

Tao Hong, Jason Wilson, and Jingrui Xie. Long-Term Probabilistic Load Forecasting and Normalization With Hourly Information. *IEEE Transactions on Smart Grid*, 5(1):456–462, Jan 2014b. ISSN 1949-3053. doi: 10.1109/TSG.2013.2274373.

Jeff Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.

Shyh-Jier Huang and Kuang-Rong Shih. Short-term load forecasting via ARMA model identification including non-gaussian process considerations. *IEEE Transactions on Power Systems*, 18(2):673–679, May 2003. ISSN 0885-8950.

Samuel Humeau, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, pages 1–6, Oct 2013. doi: 10.1109/SustainIT.2013.6685208.

O. Hyde and P.F. Hodnett. An adaptable automated procedure for short-term electricity load forecasting. *IEEE Transactions on Power Systems*, 12(1):84–94, Feb 1997. ISSN 0885-8950. doi: 10.1109/59.574927.

Rob J. Hyndman and Shu Fan. Density Forecasting for Long-Term Peak Electricity Demand. *IEEE Transactions on Power Systems*, 25(2):1142–1153, May 2010. ISSN 0885-8950. doi: 10.1109/TPWRS.2009.2036017.

Rob J. Hyndman and Yeasmin Khandakar. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1–22, 7 2008. ISSN 1548-7660.

International Energy Agency. Key world energy statistics, 2014. URL http://www.iea.org/publications/freepublications/publication/KeyWorld2014.pdf.

Ali Ipakchi and Farrokh Albuyeh. Grid of the future. *IEEE Power and Energy Magazine*, 7(2): 52–62, March 2009. ISSN 1540-7977. doi: 10.1109/MPE.2008.931384.

Aylin Jarrah Nezhad, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. Smartd: Smart meter data analytics dashboard. In *Proceedings of the 5th International Conference on Future Energy Systems*, e-Energy '14, pages 213–214, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2819-7. doi: 10.1145/2602044.2602046.

Jeff St. John. Big data on the smart grid: 2013 in review and 2014 outlook. Greentech Media, December 2013. URL http://www.greentechmedia.com/articles/read/Big-Datas-5-Big-Steps-to-Smart-Grid-Growth-in-2014.

Hayley E. Jones and David J. Spiegelhalter. Improved Probabilistic Prediction of Healthcare Performance Indicators using Bidirectional Smoothing Models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(3):729–747, 2012. ISSN 1467-985X. doi: 10.1111/j.1467-985X.2011.01019.x.

Sachin Kamboj, Willett Kempton, and Keith S. Decker. Deploying power grid-integrated electric vehicles as a multi-agent system. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '11, pages 13–20, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 0-9826571-5-3, 978-0-9826571-5-7.

KEMA, Inc. PJM Empirical Analysis of Demand Response Baseline Methods. White Paper, April 2011.

Wolfgang Ketter, John Collins, Prashant P Reddy, and Mathijs De Weerdt. The 2014 power trading agent competition. *ERIM Report Series Reference No. ERS-2013-006-LIS*, March 2014. doi: 10.2139/ssrn.2411847.

Alireza Khotanzad, Reza Afkhami-Rohani, Tsun-Liang Lu, Alireza Abaye, Malcolm Davis, and Dominic J. Maratukulam. ANNSTLF-A Neural-Network-based Electric Load Forecasting System. *IEEE Transactions on Neural Networks*, 8(4):835–846, Jul 1997. ISSN 1045-9227.

Masashi Kitayama, Ryunosuke Matsubara, and Yoshiu Izui. Application of data mining to customer profile analysis in the power electric industry. In *IEEE Power Engineering Society Winter Meeting, 2002*, volume 1, pages 632–634 vol.1, 2002. doi: 10.1109/PESW.2002.985078.

Roger Koenker and Jr. Bassett, Gilbert. Regression quantiles. *Econometrica*, 46(1):pp. 33–50, 1978. ISSN 00129682.

Koen Kok. *The PowerMatcher: Smart Coordination for the Smart Electricity Grid*. PhD thesis, Vrije Universiteit, Amsterdam, Germany, May 2013.

J. Zico Kolter and Joseph Ferreira. A large-scale study on predicting and contextualizing building energy usage. In Wolfram Burgard and Dan Roth, editors, *AAAI*. AAAI Press, 2011.

J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011.

Ramachandra Kota, Georgios Chalkiadakis, Valentin Robu, Alex Rogers, and Nicholas R. Jennings. Cooperatives for demand side management. In *The Seventh Conference on Prestigious Applications of Intelligent Systems (PAIS @ ECAI)*, pages 969–974, August 2012.

Michael Kremer. Population Growth and Technological Change: One Million B.C. to 1990. *The Quarterly Journal of Economics*, 108(3):681–716, 1993. ISSN 00335533.

Jean Kumagai. Hydrostor wants to stash energy in underwater bags. IEEE Spectrum, July 2014. URL http://spectrum.ieee.org/energy/renewables/hydrostor-wants-to-stash-energy-in-underwater-bags.

Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1):420–430, Jan 2014. ISSN 1949-3053. doi: 10.1109/TSG.2013.2278477.

Alex Laskey. How behavioral science can lower your energy bill. TED, February 2013. URL http://www.ted.com/talks/alex_laskey_how_behavioral_science_can_lower_your_energy_bill.html. Video file.

Jean-Yves Le Boudec and Dan-Cristian Tomozei. Demand response using service curves. In *IEEE PES Innovative Smart Grid Technologies*, 2011.

Jean-Yves Le Boudec and Dan-Cristian Tomozei. A demand-response calculus with perfect batteries. In *Proceedings of the 16th International GI/ITG Conference on Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance*, MMB'12/DFT'12, pages 273–287, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-28539-4.

Sylvain Lemofouet. Compressed air: the battery of the future? TheNational UAE, June 2014. URL http://www.thenational.ae/uae/uae-topics/compressed-air-the-battery-of-the-future.

Na Li, Lijun Chen, and Steven H. Low. Optimal demand response based on utility maximization in power networks. In *Proceedings of the 2011 IEEE Power & Energy Society General Meeting*, July 2011.

Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Min. Knowl. Discov.*, 15(2):107–144, October 2007. ISSN 1384-5810.

Xiaoyan Liu, Zhenjiang Lin, and Huaiqing Wang. Novel Online Methods for Time Series Segmentation. *IEEE Trans. Knowl. Data Eng.*, 20(12):1616–1626, 2008.

Zhongjing Ma, Duncan S. Callaway, and Ian A. Hiskens. Decentralized charging control of large populations of plug-in electric vehicles. *IEEE Transactions on Control Systems Technology*, 21(1):67–78, 2013. ISSN 1063-6536. doi: 10.1109/TCST.2011.2174059.

Jennifer Mankoff, Susan R. Fussell, Tawanna Dillahunt, Rachel Glaves, Catherine Grevet, Michael Johnson, Deanna Matthews, H. Scott Matthews, Robert McGuire, Robert Thompson, Aubrey Shick, and Leslie D. Setlock. Stepgreen.org: Increasing Energy Saving Behaviors via Social Networks. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, May 2010.

Francisco Martínez-Álvarez, Alicia Troncoso, José C. Riquelme, and Jesús S. Aguilar Ruiz. Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8):1230–1243, 2011. ISSN 1041-4347. doi: 10.1109/TKDE.2010.227.

Johanna L. Mathieu, Duncan S. Callaway, and Sila Kiliccote. Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing. In *The 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 4332–4339, 2011. doi: 10.1109/CDC.2011.6160628.

Aditya Mishra, David Irwin, Prashant Shenoy, Jim Kurose, and Ting Zhu. Smartcharge: Cutting the electricity bill in smart homes with energy storage. In *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, e-Energy '12, pages 29:1–29:10, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1055-0. doi: 10.1145/2208828.2208857.

Michel Misiti, Yves Misiti, Georges Oppenheim, and Jean-Michel Poggi. Optimized clusters for disaggregated electricity load forecasting. *REVSTAT*, 8:105–124, 2010.

Zaid Mohamed and Pat Bodger. Forecasting electricity consumption in New Zealand using economic and demographic variables. *Energy*, 30(10):1833 – 1843, 2005. ISSN 0360-5442.

Alaa Mohd, Egon Ortjohann, Andreas Schmelter, Nedzad Hamsic, and Danny Morton. Challenges in integrating distributed energy storage systems into future Smart Grid. In *IEEE International Symposium on Industrial Electronics*, pages 1627–1632, June 2008.

Amir-Hamed Mohsenian-Rad, Vincent W.S. Wong, Juri Jatskevich, Robert Schober, and Alberto Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Transactions on Smart Grid*, 1(3): 320 –331, Dec. 2010.

Jacob M. Montgomery, Florian M. Hollenbach, and Michael D. Ward. Ensemble Predictions of the 2012 US Presidential Election. *PS: Political Science & Politics*, 45:651–654, 10 2012. ISSN 1537-5935. doi: 10.1017/S1049096512000959.

Khosrow Moslehi and Ranjit Kumar. A reliability perspective of the Smart Grid. *IEEE Transactions on Smart Grid*, 1(1):57–64, June 2010. ISSN 1949-3053.

Steven J. Moss, M. Cubed, and Kerry Fleisher. Market segmentation and energy efficiency program design. *California Institute for Energy and Environment Berkeley*, 2008.

Raphael Nedellec, Jairo Cugliari, and Yannig Goude. GEFCom2012: Electric load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, 30(2):375 – 381, 2014. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2013.07.004.

Bijay Neupane, Torben Bach Pedersen, and Bo Thiesson. Towards flexibility detection in device-level energy consumption. In Wei Lee Woon, Zeyar Aung, and Stuart Madnick, editors, *Data Analytics for Renewable Energy Integration*, volume 8817 of *Lecture Notes in Computer Science*, pages 1–16. Springer International Publishing, 2014. ISBN 978-3-319-13289-1. doi: 10.1007/978-3-319-13290-7_1.

Hyung Seon Oh and Robert J. Thomas. Demand-side bidding agents: Modeling and simulation. *IEEE Transactions on Power Systems*, 23(3):1050 –1056, Aug. 2008.

Oracle. Smart Metering for Electric and Gas Utilities. http://www.oracle.com/us/industries/utilities/046593.pdf, 2011.

Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, 7(3):381–388, Aug 2011. ISSN 1551-3203. doi: 10.1109/TII.2011.2158841.

T. N. Palmer, G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher. Representing Model Uncertainty in Weather and Climate Prediction. *Annual Review of Earth and Planetary Sciences*, 33(1):163–193, 2005. doi: 10.1146/annurev.earth.33.092203.122552.

S. Chenthur Pandian, K. Duraiswamy, C. Christober Asir Rajan, and N. Kanagaraj. Fuzzy approach for short term load forecasting. *Electric Power Systems Research*, 76(6-7):541 – 548, 2006. ISSN 0378-7796.

Hsiao-Tien Pao. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. *Energy*, 31(12):2129 – 2141, 2006. ISSN 0360-5442.

Alex D. Papalexopoulos and Timothy C. Hesterberg. A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, 5(4):1535–1547, Nov 1990. ISSN 0885-8950.

Ken Peattie. Golden goose or wild goose? The hunt for the green consumer. *Business Strategy and the Environment*, 10(4):187–199, 2001.

Marc Pedersen. Segmenting residential customers: Energy and conservation behaviors. *ACEEE Summer Study on Energy Efficiency in Buildings*, 2008.

Petromil Petkov, Felix Köbler, Marcus Foth, Richard Medland, and Helmut Krcmar. Engaging energy saving through motivation-specific social comparison. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1945–1950, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0268-5.

B. D. Pitt and D. S. Kirschen. Application of data mining techniques to load profiling. In *Power Industry Computer Applications, 1999. PICA '99. Proceedings of the 21st 1999 IEEE International Conference*, pages 131–136, Jul 1999. doi: 10.1109/PICA.1999.779395.

Sarvapali D. Ramchurn, Perukrishnen Vytelingum, Alex Rogers, and Nicholas R. Jennings. Agent-based control for decentralised demand side management in the smart grid. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '11, pages 5–12, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems.

Sarvapali D. Ramchurn, Perukrishnen Vytelingum, Alex Rogers, and Nicholas R. Jennings. Putting the 'smarts' into the smart grid: A grand challenge for artificial intelligence. *Commun. ACM*, 55(4):86–97, April 2012. ISSN 0001-0782.

Sérgio Ramos and Zita Vale. Data mining techniques to support the classification of mv electricity customers. In *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–7, July 2008. doi: 10.1109/PES.2008.4596669.

Sérgio Ramos, Zita Vale, João Santana, and Jorge Duarte. Data mining contributions to characterize MV consumers and to improve the suppliers-consumers settlements. In *IEEE Power Engineering Society General Meeting*, pages 1–8, 2007. doi: 10.1109/PES.2007.385996.

William M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):pp. 846–850, 1971. ISSN 01621459.

P. S. Nagendra Rao. Combined Heat and Power Economic Dispatch: A Direct Solution. *Electric Power Components and Systems*, 34(9):1043–1056, 2006. doi: 10.1080/15325000600596775.

Teemu Räsänen and Mikko Kolehmainen. Feature-based clustering for electricity use time series data. In *Proceedings of the 9th International Conference on Adaptive and Natural Computing Algorithms*, ICANNGA'09, pages 401–412, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 3-642-04920-6, 978-3-642-04920-0.

Alan Rose. How big data is about to ignite smart grids worldwide. GreenBiz, October 2014. URL http://www.greenbiz.com/blog/2014/08/08/big-data-transform-smart-grids-worldwide.

Max Roser. CO2 emissions. Our World in Data, December 2014a. URL http://www.ourworldindata.org/data/resources-energy/co2-emissions/.

Max Roser. Energy production & changing energy sources. Our World in Data, December 2014b. URL http://www.ourworldindata.org/data/resources-energy/energy-production-and-changing-energy-sources/.

Max Roser. World population growth. Our World in Data, December 2014c. URL http://www.ourworldindata.org/data/population-growth-vital-statistics/world-population-growth/.

Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0):53 – 65, 1987. ISSN 0377-0427. doi: http://dx.doi.org/10.1016/0377-0427(87)90125-7.

Slobodan Ružić, Aca Vučković, and Nikola Nikolić. Weather sensitive method for short term load forecasting in electric power utility of Serbia. *IEEE Transactions on Power Systems*, 18 (4):1581–1586, Nov 2003. ISSN 0885-8950. doi: 10.1109/TPWRS.2003.811172.

Samer Saab, Elie Badr, and George Nasr. Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. *Energy*, 26(1):1 – 14, 2001. ISSN 0360-5442.

Thomas F. Sanquist, Heather Orr, Bin Shui, and Alvah C. Bittner. Lifestyle factors in U.S. residential electricity consumption. *Energy Policy*, 42(0):354 – 364, 2012. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2011.11.092.

Nicholas I. Sapankevych and Ravi Sankar. Time series prediction using Support Vector Machines: A survey. *IEEE Computational Intelligence Magazine*, 4(2):24 –38, May 2009. ISSN 1556-603X.

Sabine Schnittger and Stephen Beare. Economic implications of the proposed demand response mechanism. Technical report, Australian Energy Market Commission, 2012.

Mike Scott. Ge taps into the coolest energy storage technology around. Forbes, 2014. URL http://www.forbes.com/sites/mikescott/2014/03/21/ge-taps-into-the-coolest-energy-storage-technology-around/.

Wen Shen, Vahan Babushkin, Zeyar Aung, and Wei Lee Woon. An ensemble model for day-ahead electricity demand time series forecasting. In *Proceedings of the Fourth International Conference on Future Energy Systems*, e-Energy '13, pages 51–62, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2052-8. doi: 10.1145/2487166.2487173.

Jin Shieh and Eamonn Keogh. iSAX: Indexing and mining terabyte sized time series. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 623–631, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4.

Alex J. Smola and Bernhard Schölkopf. A tutorial on Support Vector Regression. *Statistics and Computing*, 14(3):199–222, August 2004. ISSN 0960-3174.

Eric Sortomme, Mohammad M. Hindi, S.D James MacPherson, and S.S. Venkata. Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses. *IEEE Transactions on Smart Grid*, 2(1):198–205, 2011. ISSN 1949-3053. doi: 10.1109/TSG.2010.2090913.

Goran Strbac. Demand side management: Benefits and challenges. *Energy Policy*, 36(12):4419 – 4426, 2008. ISSN 0301-4215.

Bernadette Sütterlin, Thomas A. Brunner, and Michael Siegrist. Who puts the most energy into energy conservation? A segmentation of energy consumers based on energy-related behavioral characteristics. *Energy Policy*, 39(12):8137 – 8152, 2011. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2011.10.008.

Souhaib Ben Taieb and Rob J. Hyndman. A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting*, 30(2):382 – 394, 2014. ISSN 0169-2070.

Ichiro Takeuchi, Quoc V. Le, Timothy D. Sears, and Alexander J. Smola. Nonparametric quantile estimation. *The Journal of Machine Learning Research*, 7:1231–1264, December 2006. ISSN 1532-4435.

James W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152, 2010.

The Commission for Energy Regulation (CER) and the ISSDA. Smart Metering Electricity Customer Behaviour Trials. Accessed via the Irish Social Science Data Archive - www.ucd.ie/issda, 2012.

Axel Tidemann, Boye A. Høverstad, Helge Langseth, and Pinar Öztürk. Effects of scale on load prediction algorithms. In *22nd International Conference on Electricity Distribution*, 2013.

George J. Tsekouras, Nikos D. Hatziargyriou, and Evangelos N. Dialynas. Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Transactions on Power Systems*, 22(3):1120–1128, Aug 2007. ISSN 0885-8950. doi: 10.1109/TPWRS.2007.901287.

United Nations. World Population Prospects: The 2012 Revision. Department of Economic and Social Affairs, December 2014. URL http://esa.un.org/unpd/wpp/unpp/panel_population.htm.

U.S. Census Bureau. Total midyear population for the world: 1950-2050, December 2014. URL https://www.census.gov/population/international/data/worldpop/table_population.php.

U.S. Department of Energy. Grid 2030: A national vision for electricity's second 100 years, July 2003.

U.S. Department of Energy. Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them, February 2006.

U.S. Environmental Protection Agency. Atmospheric concentrations of greenhouse gases, May 2014. URL http://www.epa.gov/climatechange/science/indicators/ghg/ghg-concentrations.html.

Emmanouil Valsomatzis, Katja Hose, and Torben Bach Pedersen. Balancing energy flexibilities through aggregation. In Wei Lee Woon, Zeyar Aung, and Stuart Madnick, editors, *Data Analytics for Renewable Energy Integration*, volume 8817 of *Lecture Notes in Computer Science*, pages 17–37. Springer International Publishing, 2014. ISBN 978-3-319-13289-1. doi: 10.1007/978-3-319-13290-7_2.

Stijn Vandael, Bert Claessens, Maarten Hommelberg, Tom Holvoet, and Geert Deconinck. A scalable three-step approach for demand side management of plug-in hybrid vehicles. *IEEE Transactions on Smart Grid*, 4(2):720–728, June 2013. ISSN 1949-3053. doi: 10.1109/TSG.2012.2213847.

Matteo Vasirani and Sascha Ossowski. Smart consumer load balancing: State of the art and an empirical evaluation in the Spanish electricity market. *Artificial Intelligence Review*, 39(1):81–95, 2013. ISSN 0269-2821. doi: 10.1007/s10462-012-9391-6.

Matteo Vasirani, Tri Kurniawan Wijaya, Georgios Liassas, and Karl Aberer. Privacy Enhanced Demand Response with Reputation-based Incentive Distribution. In *International Workshop on Demand Response*, Cambridge, UK, 2014.

Andreas Veit, Christoph Goebel, Rohit Tidke, Christoph Doblander, and Hans-Arno Jacobsen. Household electricity demand forecasting: Benchmarking state-of-the-art methods. In *International Workshop on Demand Response*, 2014.

Laurynas Šikšnys, Mohamed E. Khalefa, and Torben Bach Pedersen. Aggregating and disaggregating flexibility objects. In Anastasia Ailamaki and Shawn Bowers, editors, *Scientific and Statistical Database Management*, volume 7338 of *Lecture Notes in Computer Science*, pages 379–396. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-31234-2. doi: 10.1007/978-3-642-31235-9_25.

Perukrishnen Vytelingum, Thomas D. Voice, Sarvapali D. Ramchurn, Alex Rogers, and Nicholas R. Jennings. Agent-based micro-storage management for the smart grid. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*, AAMAS '10, pages 39–46, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 978-0-9826571-1-9.

Perukrishnen Vytelingum, Thomas D. Voice, Sarvapali D. Ramchurn, Alex Rogers, and Nicholas R. Jennings. Theoretical and practical foundations of large-scale agent-based micro-storage in the smart grid. *J. Artif. Intell. Res. (JAIR)*, 42:765–813, 2011.

Thomas T. Warner. *Numerical Weather and Climate Prediction*, volume 89. Cambridge University Press Cambridge, UK, 2011.

Markus Weiss, Adrian Helfenstein, Friedemann Mattern, and Thorsten Staake. Leveraging smart meter data to recognize home appliances. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, pages 190 –197, March 2012.

Rafal Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. HSC Research Reports HSC/14/07, Hugo Steinhaus Center, Wroclaw University of Technology, May 2014.

Tri Kurniawan Wijaya, Dipyaman Banerjee, Tanuja Ganu, Dipanjan Chakraborty, Sourav Battacharya, Thanasis G. Papaioannou, Deva P. Seetharam, and Karl Aberer. DRSim: A cyber physical simulator for demand response systems. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 217–222, Oct 2013a. doi: 10.1109/SmartGridComm.2013.6687960.

Tri Kurniawan Wijaya, Julien Eberle, and Karl Aberer. Symbolic representation of smart meter data. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 242–248. ACM, 2013b. ISBN 978-1-4503-1599-9.

Tri Kurniawan Wijaya, Kate Larson, and Karl Aberer. Matching demand with supply in the smart grid using agent-based multiunit auction. In *The Fifth International Conference on Communication Systems and Networks (COMSNETS)*, pages 1–6, 2013c.

Tri Kurniawan Wijaya, Thanasis G. Papaioannou, Xin Liu, and Karl Aberer. Effective consumption scheduling for demand-side management in the smart grid using non-uniform participation rate. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, pages 1–8, Oct 2013d. doi: 10.1109/SustainIT.2013.6685188.

Tri Kurniawan Wijaya, Tanuja Ganu, Dipanjan Chakraborty, Karl Aberer, and Deva P. Seetharam. Consumer segmentation and knowledge extraction from smart meter and survey data. In Mohammed Javeed Zaki, Zoran Obradovic, Pang-Ning Tan, Arindam Banerjee, Chandrika Kamath, and Srinivasan Parthasarathy, editors, *Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, Pennsylvania, USA*, pages 226–234. SIAM, April 2014a. ISBN 978-1-61197-344-0. doi: 10.1137/1.9781611973440.26.

Tri Kurniawan Wijaya, Samuel Franois Roger Joseph Humeau, Matteo Vasirani, and Karl Aberer. Individual, Aggregate, and Cluster-based Aggregate Forecasting of Residential Demand. Technical report, EPFL, Lausanne, Switzerland, 2014b.

Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. Crowdsourcing Behavioral Incentives for Pervasive Demand Response. Technical report, EPFL, Lausanne, Switzerland, 2014c.

Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. When bias matters: An economic assessment of demand response baselines for residential customers. *IEEE Transactions on Smart Grid*, 5(4):1755–1763, July 2014d. ISSN 1949-3053.

Tri Kurniawan Wijaya, Matteo Vasirani, Jonas Christoffer Villumsen, and Karl Aberer. Characterizing the Optimal Incentives for Pervasive Demand Response. Technical report, EPFL, Lausanne, Switzerland, 2014e.

Tri Kurniawan Wijaya, Mathieu Sinn, and Bei Chen. Forecasting Uncertainty in Electricity Demand. In *AAAI-15 Workshop on Computational Sustainability*, 2015.

Simon N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2006.

World Commission on Environment and Development. *Our common future*. Oxford University Press, Oxford, 1987. ISBN 019282080X.

Matt Wytock and J. Zico Kolter. Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields. In *IEEE 52nd Annual Conference on Decision and Control (CDC)*, pages 1019–1024, Dec 2013. doi: 10.1109/CDC.2013.6760016.

Curriculum Vitae

Personal Data

NAME:	Tri Kurniawan WIJAYA
PLACE AND DATE OF BIRTH:	Indonesia — 07 June 1985
ADDRESS:	EPFL IC LSIR BC147, Station 14 CH-1015 Lausanne Switzerland
PHONE:	+41 21 69 32681
EMAIL:	`tri-kurniawan.wijaya@epfl.ch`

Education

SINCE OCT 2011 | PhD student at the DISTRIBUTED INFORMATION SYSTEM LAB, School of Computer and Communication Sciences, **EPFL**, Switzerland.
Currently also working on EU FP7 Wattalyst project, smart meter data analytics, and pervasive demand response systems. | www.wattalyst.org
Supervisor: Prof. Karl ABERER.
Expected graduation date: APRIL 2015

OCT 2009–SEPT 2011 | Master of COMPUTATIONAL LOGIC ,

- International Center for Computational Logic. Faculty of Computer Science, **Dresden University of Technology**, Germany, and
- Faculty of Informatics, **Vienna University of Technology**, Austria.

Graduated with ***distinction***
Thesis: "Top Down Evaluation Techniques for Modular Nonmonotonic Logic Programs." | www.trikurniawanwijaya.com/master-thesis
Supervisor: Prof. Thomas EITER.
GPA: 1.2 (min=1.0, max=5.0, the lower the better)

AUG 2003-OCT 2007 | Undergraduate Degree in INFORMATION SYSTEMS AND COMPUTER ENGINEERING, **Sekolah Tinggi Teknik Surabaya**, Indonesia.

	Graduated with ***Magna Cum Laude*** Thesis: "Implementation of WaveCluster, STING and GRIDCLUS Algorithms in Astronomy Data Clustering" Supervisor: Prof. Kuswara Setiawan. GPA: 3.85 (min=0.0, max=4.0, the higher the better)
Jul 2000–Jun 2003	High School, **SMU Negeri 1 Genteng**, Banyuwangi, Jawa Timur, Indonesia.

Awards and Scholarships

Mar 2013	IBM Ph.D. Fellowship Award
Oct 2010	2^{nd} place, German General Game Playing Competition, Potsdam University, Germany
Mar 2010	2^{nd} place, Dresden General Game Playing Competition, Dresden University of Technology, Germany
Oct 2009–Sep 2011	Erasmus Mundus Master Course Scholarship, European Commission
Aug 2009	1^{st} place, Indonesian Robotic Olympiad (team coach)
Jul 2008	1^{st} place, Indonesian Robotic Olympiad (team coach)
Mar 2008	Merit scholarship, Monash University, Australia
Jul 2007	1^{st} place, Indonesian Robotic Olympiad (team coach)
Mar 2005	1^{st} place, Petra Informatics Competition, Petra Christian University, Indonesia
Feb 2004–Feb 2007	PRESTASI scholarship (GPA $\geq 3.8/4.0$), Sekolah Tinggi Teknik Surabaya, Indonesia
2002	1^{st} Place, Java-Bali Physics Contest, Institut Teknologi Sepuluh Nopember, Indonesia
Dec 2001	High Distinction, Australian National Chemistry Quiz Junior Division, Australia

Teaching

Sep 2014–Jan 2015	Teaching Assistant. École Politechnique Fédérale de Lausanne (EPFL), Switzerland. Analysis (Calculus) I
Sep 2012–Jan 2013	Teaching Assistant. École Politechnique Fédérale de Lausanne (EPFL), Switzerland. Distributed Information Systems
Dec 2007–Sep 2009	Lecturer. Sekolah Tinggi Teknik Surabaya, Indonesia. Teaching and training, Information Systems and Computer Engineering faculty.
Feb 2007–Jun 2007	Teaching Assistant. Sekolah Tinggi Teknik Surabaya, Indonesia. Visual Programming, Database, and Computer Networking
Aug 2005–Nov 2007	Full-time Computer Laborator Assistant. Sekolah Tinggi Teknik Surabaya, Indonesia.

	Pascal Programming, C/C++ Programming, Java Programming, Delphi Programming, Visual Basic .NET programming
Aug 2004–Oct 2004	Part-time Computer Laborator Assistant. Sekolah Tinggi Teknik Surabaya, Indonesia. Help students to understand their current laboratory material, assess students' exercises.

Internships

Jan-May 2014	Exploratory Predictive Analtyics, **IBM Research**, Ireland Quantify and forecast uncertainty in electricity demand. Carefully analyze electricity consumption of France (national level) and Ireland (residential level). Build relationships between electricity consumption and various external factors, such as weather/temperature, holiday, day types (weekday/weekend), and hour of the day. Technical skills: R and Java.
Jun-Aug 2013	Smarter Energy, **IBM Research**, India Develop context-based demand response baselines and consumer segmentation (clustering) framework, discover consumer cluster changes over time, and identify main characteristics of a cluster from consumer surveys. Technical skills: Java, Weka, MySQL, and gnuplot.

Publications

Smart Meter Data Analytics

[1] J. Eberle, T. K. Wijaya and K. Aberer
Online Unsupervised State Recognition in Sensor Data
IEEE International Conference on Pervasive Computing and Communications **(PerCom)**, St. Louis, Missouri, USA, 2015.

[2] T. K. Wijaya, M. Sinn and B. Chen
Forecasting Uncertainty in Electricity Demand
AAAI-15 Workshop on Computational Sustainability, Austin, TX, USA, 2015.

[3] H.-Â. Cao, T. K. Wijaya and K. Aberer
Estimating Human Interactions with Electrical Appliances for Activity-based Energy Savings Recommendations (poster)
1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings **(BuildSys)**, Memphis, TN, USA, 2014.

[4] A. Jarrah Nezhad, T. K. Wijaya, M. Vasirani and K. Aberer
SmartD: Smart Meter Data Analytics Dashboard (demo)
The 5th ACM International Conference on Future Energy Systems **(e-Energy)**, Cambridge, UK, 2014.

[5] T. K. Wijaya, T. Ganu, D. Chakraborty, K. Aberer and D. P. Seetharam
Consumer Segmentation and Knowledge Extraction from Smart Meter and Survey Data

SIAM International Conference on Data Mining **(SDM)**, Philadelphia, Pennsylvania, USA, 2014.

[6] S. Humeau, T. K. Wijaya, M. Vasirani and K. Aberer
Electricity Load Forecasting for Residential Customers: Exploiting Aggregation and Correlation between Households
Sustainable Internet and ICT for Sustainability **(SustainIT)**, Palermo, Italy, 2013.

[7] T. K. Wijaya, J. Eberle and K. Aberer
Symbolic Representation of Smart Meter Data
Proceedings of the Joint **EDBT/ICDT 2013 Workshops**, p. 242-248. ACM, New York, NY, USA, 2013.

Pervasive Demand Response

[1] V. Chandan, T. Ganu, T. K. Wijaya, M. Minou, G. Stamoulis, G. Thanos, and D. P. Seetharam
iDR: Consumer and Grid Friendly Demand Response System
The 5th ACM International Conference on Future Energy Systems **(e-Energy)**, Cambridge, UK, 2014.

[2] M. Vasirani, T. K. Wijaya, G. Liassas and K. Aberer
Privacy Enhanced Demand Response with Reputation-based Incentive Distribution
International Workshop on Demand Response, in conjunction with the 5th ACM International Conference on Future Energy Systems (e-Energy), Cambridge, UK, 2014.

[3] T. K. Wijaya, M. Vasirani and K. Aberer
When Bias Matters: An Economic Assessment of Demand Response Baselines for Residential Customers
IEEE Transactions on Smart Grid, 2014.

[4] T. K. Wijaya, T. G. Papioannou, X. Liu and K. Aberer
Effective Consumption Scheduling for Demand-Side Management in the Smart Grid using Non-Uniform Participation Rate
Sustainable Internet and ICT for Sustainability **(SustainIT)**, Palermo, Italy, 2013.

[5] T. K. Wijaya, D. Banerjee, T. Ganu, D. Chakraborty and S. Battacharya, T. G. Papaionannou, D. P. Seetharam, and K. Aberer
DRSim: A Cyber Physical Simulator for Demand Response Systems
IEEE International Conference on Smart Grid Communications **(SmartGridComm)**, Vancouver, Canada, 2013.

[6] T. K. Wijaya, K. Larson and K. Aberer
Matching Demand with Supply in the Smart Grid using Agent-Based Multiunit Auction
The 5th International Conference on Communication Systems and Networks (**COMSNETS**), Bangalore, India, 2013.

Declarative Problem Solving

[1] Q. V. H. Nguyen, T. T. Nguyen, V. T. Chau, T. K. Wijaya, Z. Miklos, K. Aberer, A. Gal and M. Weidlich

SMART: A tool for analyzing and reconciling schema matching networks. (demo)
31st IEEE International Conference on Data Engineering **(ICDE)**, Seoul, Korea, 2015.

[2] Q. V. H. Nguyen, T. K. Wijaya, Z. Miklos, K. Aberer, E. Levy, V. Shafran, A. Gal and M. Weidlich
Minimizing Human Effort in Reconciling Match Networks
The 32nd International Conference on Conceptual Modeling **(ER)**, Hong Kong, 2013.

Activity Recognition

[1] A. Rai, Z. Yan, D. Chakraborty, T. K. Wijaya and K. Aberer
Mining Complex Activities in the Wild via a Single Smartphone Accelerometer
The 6th International Workshop on Knowledge Discovery from Sensor Data, **(SensorKDD)**, Beijing, China, 2012.

Reviewing

Conference	VLDB (2012), ICDE (2013, 2015), CIKM (2014), ICDCS (2012), WISE (2012) BPM (2012), P2P (2012)
Journal	IEEE Transactions on Smart Grid 2014, Elsevier COMCOM 2014

Work Experience

May 2010–Sep 2010	Part-time working student, SAP Research AG, Dresden, Germany. Worked in an already running, large project, Aletheia (Semantic federation of comprehensive product information). Designed and implemented context-based services with the Sesame Framework. Developed the authentication module and the context-based Graphical User Interface to retrieve product information.
Feb 2009–Sep 2009	Private Assistant of the Associate Dean I, Sekolah Tinggi Teknik Surabaya, Indonesia. Provided support for organizational and administrative tasks of the Associate Dean I.
Aug 2006–Aug 2009	Robotics Teaching Staff, RoboKidz Computer Learning Center, Indonesia. Trained students to create and program Lego Mindstorms Robotic and prepared them to compete in the National championship.
Aug 2005–Nov 2007	Full-time Assistant at the Computer Laboratory, Sekolah Tinggi Teknik Surabaya, Indonesia Maintained hardware and software, designed and prepared laboratory activity, helped students to understand their current laboratory material, and graded students' laboratory exercise.

Organizational Experience

Jun 2011	Erasmus Mundus Course Representatives for the European Master's Program in Computational Logic at the General Assembly of Erasmus Mundus Association, Budapest, Hungary
Apr 2011	Student representative for the Joint Committee meeting at EMCL Workshop 2011 Lisbon, Portugal
Apr–Sep 2010	Study and Examination Board of the International Master Program of Computational Logic, Faculty of Computer Science, Dresden University of Technology, Germany
Feb 2008 & 2009	Chief of the Judge Committee of the Regional Robot Olympiad, Surabaya, Indonesia
Nov 2006	Coordinator of General Affairs of the 23rd Graduation Ceremony, Sekolah Tinggi Teknik Surabaya, Indonesia
Aug 2004–Aug 2005	Head of Secretarial Bureau of the Student Senate, Sekolah Tinggi Teknik Surabaya, Indonesia
2003	Co-founder and Chairman of the Physics Club, SMU Negeri 1 Genteng, Banyuwangi, Indonesia
2002	Co-founder and Vice Chairman of the Mathematics League for Junior High School in the Banyuwangi Regency, Indonesia

Language Skills

Indonesian:	Mother tongue	
English:	C1	Common European Framework of Reference for Languages
German:	A2	
French:	A1	

Personal Interests and Activities

Reading: Politics, Economics, Management (learn new things)
Movies, Music (creativity)
Travelling, Photography (adventure and inspiration)
Football, Badminton (fun)