DTU

Integration of fluctuating energy by electricity price control

Master Thesis Olivier Corradi & Henning Ochsenfeld 04.08.2011

DTU Informatics
Department of Informatics and Mathematical Modelling

DONG energy

Preface

The energy system stands at a crossroad, as electricity consumption and production in Denmark is set to change significantly in the coming years. Electricity generation will increasingly rely on wind energy, and consumption will adapt accordingly by increasing its dependency on electricity rather than fossil fuels. Plug-in hybrid electric vehicles are already appearing on the market, imposing high constraints on the electricity grid. Ted Craver, chairman and CEO of the Californian utility Edison International, dares to go as far as saying that „there will be more changes in the electricity industry in the next 10 years than there were in the past 125 years we have been in business". IT and communication lay the ground for this paradigm shift, enabling actors to collaborate and communicate more efficiently in the future power system, often referred to as the *Smart Grid*.

Traditionally, the power sector has adapted and reinforced the power grid by laying more and thicker cables in the ground. A recent study by Energinet.dk and the Danish Energy Association concluded that the social net cost of setting up a Smart Grid is DKK 1.6 billion in contrast to the DKK 7.7 billion of a traditional expansion scheme. The value creation of a Smart Grid mainly comes from the derivative benefits it creates, representing savings in electricity generation, regulating power and reserves.

As this profound transformation occur, energy and information will have to transit together as a new entity, raising new challenges in terms of infrastructure and security. The increasing reliance on fluctuating sources of energy propagates risk and uncertainty to the whole electricity value chain, challenging existing market

structures and balancing strategies.

Information and communication technologies along with statistical modelling play a central role in forging the tools necessary to build a Smart Grid. As those tools are beginning to be designed and developed, it is truly exciting to be able to contribute with a small achievement on the way to solve one of the biggest challenges of tomorrow.

Kongens Lyngby, Denmark
August 2011 *Olivier Corradi & Henning Ochsenfeld*

Acknowledgements

This thesis and the ideas behind it would not have seen the light without our supervisor, Henrik Madsen, who supported us and always took the time to help us when needed.

We would like to thank Klaus Baggesen Hilger, Mathias Jesper Dahl-Sørensen and Lars Henrik Hansen from DONG Energy for their valuable collaboration and precious insights during this project.

Finally, we are grateful to everyone who contributed with ideas and solutions to this project, involving members of the department of statistics of DTU Informatics, and especially Scott Otterson who actively took part in the initial discussions.

Abstract

Integration of fluctuating energy such as wind power becomes more and more essential in future energy systems. The reliance on it propagates risk and uncertainty to the whole electricity value chain, challenging existing market structures and balancing strategies. One solution is to take advantage of the flexibility of consumers. This is done by acting on devices having a high inertia which therefore can be turned off during a short period of time without impacting user comfort. In this thesis, the heating system of households is considered, represented by price-sensitive heat pumps. By reacting to a price signal, they automatically adjust their consumption in order to minimize energy costs. Such a price-responsive population of households is investigated through a combination of a real-life experiment and a simulation framework specially designed. The price-response is modelled and extracted, and its dependencies on system parameters are examined. In the light of a real-life implementation, a non-linear forecasting model is developed, requiring as data an aggregate metering of household consumptions, the associated price signal and weather forecasts. Adaptive estimation of the forecasting model is implemented, permitting the development of a predictive controller generating prices for the simulated households. The proof-of-concept is illustrated by following a constant consumption reference, yielding a reduction in peak consumption of nearly 5% and a mean daily consumption shift of 11%.

Contents

Chapter 1

Introduction and background

Denmark has ambitious political climate and energy targets to reduce CO_2 emissions, involving a deep integration of renewable energies, especially wind power, into electricity generation. The objective is to integrate 50% wind energy by 2025. Due to its highly fluctuating nature, such a penetration of wind energy can only be achieved if a high degree of flexibility is introduced in the consumption as an efficient large-scale electricity storage solution is yet to be found. In a Smart Grid, consumers will be able to interact with the power system, introducing completely new perspectives.

This project aims at demonstrating to which extent the flexibility of households can be activated by means of a varying electricity price. We will focus on the appliances that offer the highest flexibility potential, even though the procedure may be generalized to other systems (and even to companies or industries).

Our goal is to create a *proof-of-concept*, and to investigate its potentials and limits. As no such flexibility is available in the Danish society, we have based our study on a combination of real data and on a simulation framework, specially designed

for this purpose. The concept is then tested using the simulation framework, by building a price generator steering the consumption of end users.

1.1 The Danish electricity system

The Danish electricity system can be seen as a three-layer system consisting of production, transmission and distribution/consumption. It contains two markets, both liberalized: The retail market, where end users buy electricity from an electricity supplier, and the wholesale market, where market actors trade significant amounts of electricity. The wholesale market is Nord Pool, the largest market for electrical energy in the world. The retail market was liberalized in 2003, enabling end users to chose their electricity provider, and the wholesale market was liberalized in 1999, to enhance free competition in cross–border electricity production and trade, decoupling the transmission grid from electricity generation.

1.1.1 Actors

A variety of actors are present, with very different but crucial roles (Figure 1.1).

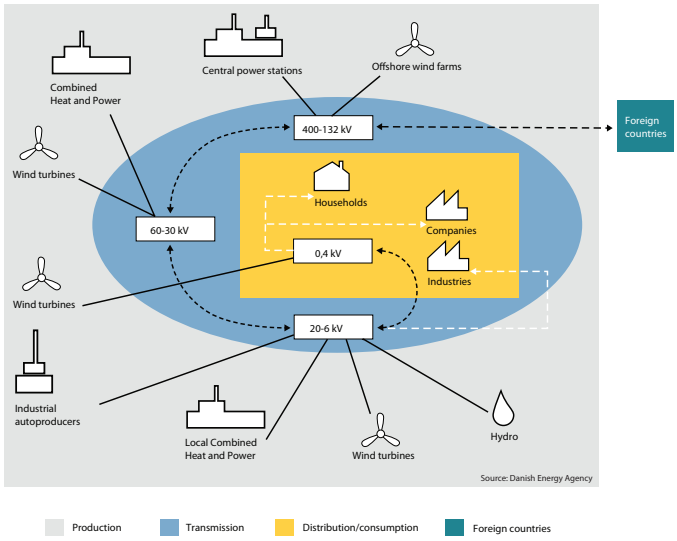


Figure 1.1: *The Danish electricity system consists of three layers: Production, transmission and distribution/consumption.*

Transmission system operator (TSO) Owns the high voltage installations (the

backbone of the electricity system) and the international connections. Its role is to transmit electrical power from generation plants to regional electricity distribution operators and to ensure stability and balance in the energy system. To do so, it accepts bids for *regulating* power on the regulating power market. The TSO is also involved in handling ancillary services maintaining grid stability. Balancing the energy system means maintaining grid frequency (related to speed of rotation of generators) by balancing the amount of generated electricity with the amount consumed. In Denmark, the TSO is represented by the state-owned monopoly *Energinet.dk*. Furthermore, the TSO develops market rules and regulations that provide a framework for a well–functioning energy market.

Distribution system operator (DSO) Operates the distribution network and meters data on production and consumption. There are multiple DSOs in Denmark, acting as monopolies in each region.

Generating companies Produces electricity and sells it either directly to an electricity supplier or to Nord Pool.

Retailers Concludes contracts with consumers about the supply of electricity. The electricity supplier buys electricity either at a power exchange market (Nord Pool), from an electricity producer or from a third party electricity trader. The end user has the right to change from one electricity supplier to another through the retail market.

Balance responsible parties (BRP) Production, consumption and trade activities are assigned to the BRP who enters an agreement with Energinet.dk, assuming responsibility for one or several specific activities (production, consumption or trading). The BRP assumes the financial responsibility for the imbalances they cause. Note that generating companies (produces) have the possibility to sell regulating power directly to Energinet.dk via the regulating power market if they act as a BRP. This is also the case for retailers.

Nord Pool - the Nordic power exchange Nord Pool is the Nordic power exchange market and is completely owned by Nordic TSOs.

The TSO and the DSOs are regulated monopolies, and are subject to strict regulation. One company can take on multiple roles, like e. g. DONG Energy who represents both a balance responsible, retailer and producer.

1.1.2 Various markets for electricity trading

Most of the trading between market actors is made via the power exchange Nord Pool, a common Nordic power exchange market owned by all Nordic TSOs (Finland, Norway, Sweden and Denmark). Nord Pool was established in 1996, and Denmark fully joined in 2000. Over 75% of all electricity consumption in the Nordic countries is traded through Nord Pool[1].

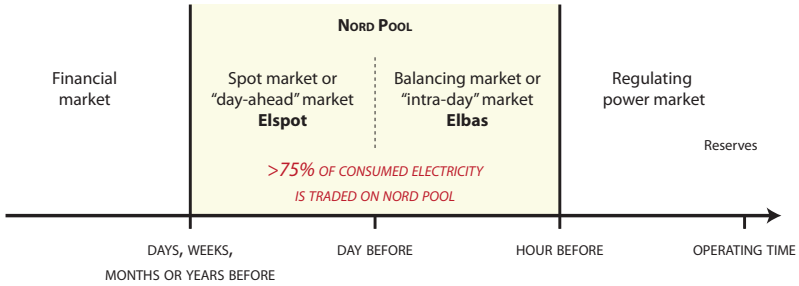


Figure 1.2: *Different markets operate at different time scales in order to adjust to unforeseen deviations from a consumption and production schedule.*

Different markets operating at different time scales provide flexibility to unforeseen events that could cause deviations in the production and consumption schedule (Figure 1.2). Those events could be infrastructure breakdowns, fluctuations in wind energy or simply a manifestation of the stochastic behaviour of consumers using energy when they see fit. The following markets are used [7]:

The day-ahead market: Elspot Electricity for delivery the next day is traded on Elspot (also referred to as Nord Pool Spot). The trade results in a price that may be characterised as the market price of electricity in the Nordic countries. Elspot is based on the auction principle, i. e. all players that want to buy or sell energy make bids consisting of a price (Euro) and quantity (MW). Bidding is permitted up to 12:00 noon before the day of operation. One hour after, Nord Pool establishes the market price (also called clearing price or system price) by matching supply and demand curves (Figure 1.3).

The intra-day market: Elbas Electricity can be traded up to one hour before the delivery hour on Elbas. This makes it possible for players to buy and sell as required in order to ensure balance right up to the delivery hour,

[1] `http://energitilsynet.dk`

for example in case of outages or deviations in the planned wind energy production.

The Nordic regulating power market Also based on the auction principle, the regulating power market permits electricity trading between the Nordic TSOs and BRPs to maintain this balance. Trading on this market is possible up till 30 minutes before the operating hour. The TSOs can either submit bids for increased production (upwards regulation) or reduced production (downwards regulation). Capacity bought or sold is to be activated in 15 minutes and kept until the end of the operating hour. The offered volumes of regulating power bids are forwarded to the Nordic Operation Information System (NOIS) which is a common Nordic list including bids from Danish, Norwegian, Swedish, and Finnish partners. If there is a need for upregulation, the Danish TSO activates (e. g. accepts bids from the BRP) the needed volume of regulating power at the lowest bid price on the NOIS list. On the other hand, in times with a need for downregulation, the TSO accepts bids for selling the surplus volume to the BRP with the highest bid for downregulation.

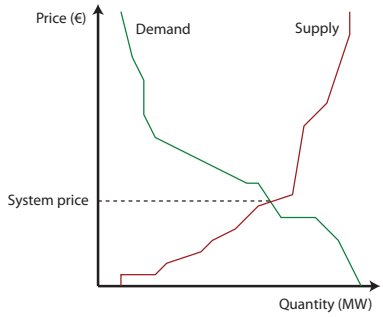


Figure 1.3: *Suppliers and consumers make bids on a market according to their electricity needs or available capacity. The aggregated demand curve is constructed by ranking bids from highest to lowest, and the equilibrium price is then found as the intersection of both curves.*

If deviations from the scheduled production and consumption produce imbalances in the system that no market can cover, the TSOs have emergency reserves that can be used to restore balance (in the case of a major breakdown for example).

1.1.3 Imbalance settlements

The grid companies have the obligation to supply the Danish TSO Energinet.dk with information on metered data for the consumption and production of each BRP. Every day, Energinet.dk sends out to the BRP their metered consumption/production for the day before. Imbalances between the schedule from the BRP and the real amount metered, are then settled financially [8].

1.2 The concept of consumption control by price

In the current setup, supply and demand are balanced using markets to activate the flexibility of production. Depending on market prices, the profitability of producers mainly depends on the type of production used. For example, hydropower is very flexible, as its resource (water) is storable. It has the advantage of being able to produce when the electricity prices are high. On the contrary, wind energy is non-storable, and its profitability therefore directly fluctuates with the wind itself and the market prices (if the operating costs remain constant). Some production means, because of their expensive nature, are only activated in situations of very high demand, where the price is high enough to cover the operating costs. As an example, diesel generators can quickly be activated – but to a high cost. Furthermore, in the case of overproduction (too much wind for example), producers can be paid to stop producing, enabling them to cover their operating expenses preventing an unnecessary production.

In that sense, the flexibility of production is based on the logic of maximizing profits by exchanging energy on the power markets. The flexibility of consumption on the other hand is not fully activated, because the underlying structure simply does not exist (both in terms of market and infrastructure). Because end-users are hidden behind the electricity supplier, the TSO is not in direct contact with the end-user, and therefore can not activate its available flexibility. However, it is most likely that end users are willing to become flexible if price incentives are provided. This flexibility of demand, or response to a certain control signal, is called *demand response*.

1.2.1 Different objectives for multiple actors

Consumers can be very different in their control objectives. Some might be very interested in reducing costs, others in reducing environmental impact or even in increasing comfort by adding intelligence in their homes (in the case of households). Industry consumers are however most likely to have an objective based on maximizing profit. Controlling the consumption of end users means following and coordinating those multiple objectives. The price incentive might be the most motivating control signal, but one could imagine to control the consumption with a mixture of electricity price and CO_2 footprint of production.

The generating companies represent a broad range of actors, ranging from a single wind turbine to large companies with a portfolio of power producing units. Its main objective is however to maximize profit, and has little interest in controlling the consumption.

The TSO is responsible for ensuring the security of supply. Its objective is therefore to balance production and consumption, such that reserves are always available when needed and such that no unused power is generated. The TSO has no direct control over production or consumption. An indirect form of control takes place through the regulating power market, where electricity prices stabilize the exchange of power, both inside Denmark and with the neighbouring countries. It could therefore be interesting for the TSO to extend the power markets to the end-users, so that they also become price-responsive, in the same way that power producers already react to prices.

The DSO is responsible for the distribution of electricity and the subsequent infrastructure used. Its objective is to provide sufficient grid access at any time and to avoid congestion in case of high load. By discriminating geographically electricity prices of end-users, the DSO could better manage congestions and even prevent breakdowns of weak parts of the system.

The BRP and the electricity supplier (or retailer) buy or sell electricity. Their objective is to maximize profits, and therefore they are interested in buying and selling at the right time. Furthermore, the BRP pays penalties for the imbalance it has caused in deviating from its planned consumption or production. Controlling the consumption with price is therefore a mean to minimize the paid penalties by adjusting the consumption on a short time scale such that it follows the submitted plan (Figure 1.4).

In the light of a proof-of-concept, we will focus on the control of consumption,

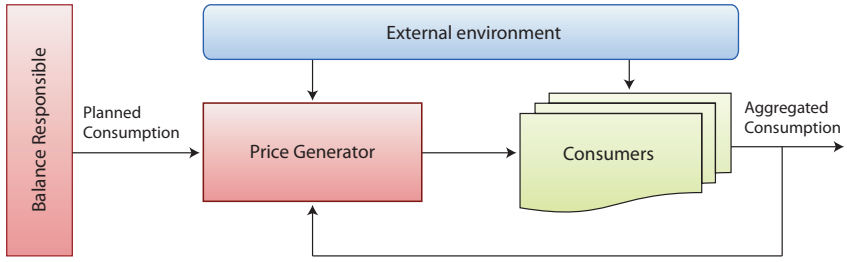


Figure 1.4: *One of the possible implementations of a control-by-price strategy would aim at minimizing the penalties paid by the Balance Responsible Party by adjusting, on a short time scale, the consumption such that it follows the original schedule.*

following the objectives of the BRP. By improving the capabilities of the BRP to follow a certain consumption schedule, higher shares of fluctuating energies can then be introduced.

Identifying the flexible part of the consumption

In order to be able to control the consumption, one must assess which part of it is actually flexible. The 2009 Danish electricity consumption can be broken down three main parts [1].

- **Commercial activities and utilities**, 11 918 GWh (37% of total)
- **Industry and agriculture**, 10 752 GWh (33% of total)
- **Households**, 9 495 GWh (30% of total)

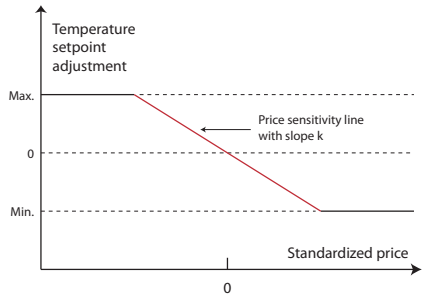
As a starting point for simplicity, households are considered, accounting for 30% of the total Danish electricity consumption. Inside Danish households, the electricity consumption is broken down into the following groups [1],

- **Fridge and Freezer**, flexible (16% of total)
- **Washing and drying**, somewhat flexible (15% of total)
- **Heating (air+water)**, flexible (14% of total)
- **TV/Hifi**, not flexible (14% of total)

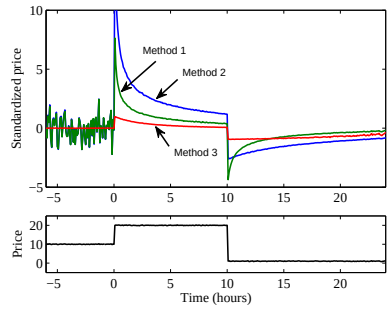
- **Light**, not flexible (13% of total)
- **Others**, not flexible (10% of total)
- **Computer**, not flexible (10% of total)
- **Cooking**, not flexible (8% of total)

where flexibility describes the extent to which the appliances in the given category are able to defer their consumption. Another way to phrase this would be to define flexible appliances as systems that have a significant inertia, or sufficiently long time constants, such that they can be turned off during a certain period of time without significantly impacting user comfort.

Heating and cooling are found to represent the biggest portion (in consumption) of price-responsive appliances. Their flexibility could be activated by lowering or increasing their temperature thermostat (setpoint) depending on the electricity price. Some hard comfort bounds can then be applied to prevent the temperature setpoints to diverge from a certain comfort zone (Figure 1.5a).



(a) *A change in price yields a change of the heating set point within a certain comfort zone.*



(b) *Price smoothing strategies with $\tau = 24$ hours. Method 3 captures a doubling of the price as an increase of +100%.*

Figure 1.5

However, in order to adjust the temperature setpoint depending on a price, the end-user has to assess how high or low the price of electricity is at a given point in time. This can only be achieved if the current price is compared to a reference, which could be a daily mean for example. This allows for slow price variations to be filtered out (such as monthly or yearly variations), as one would like to avoid

turning off appliances caused by increase of prices over one year for example. The received market price p_t can therefore be transformed into a dimensionless *standardized* price ρ_t, which is then independent of currency and insensitive to slow variations:

$$\rho_t = \frac{p_t - \mu_t}{\sigma_t}. \tag{1.1}$$

μ_t is the price reference, and σ_t a standardization factor accounting for the price volatility. A simple approach is to compute μ_t as the mean over 24 hours, and σ_t as the standard deviation over the same 24 hours [9]. Note that this first method equally weights past measurements.

[19] proposed a second method having μ_t computed as an average with exponentially decaying weights, also called exponentially weighted moving average. Nevertheless, the two previous approaches have the major drawback that when the price is close to constant, the standard deviation is close to zero, yielding very high fluctuations in the standardized price (Figure 1.5b).

We therefore propose the idea of computing the *increase* in price with respect to a certain reference $\bar{p}_t$, where the latter is the exponentially weighted moving average previously mentioned (method 3):

$$\rho_t = \frac{p_t}{\bar{p}_{t-1}} - 1. \tag{1.2}$$

The exponentially weighted moving average, in its recursive form, is stated as

$$\bar{p}_t = \bar{p}_{t-1} + \frac{\Delta t}{\Delta t + \tau} \left(p_t - \bar{p}_{t-1} \right), \tag{1.3}$$

where τ is the time constant accounting for how long a price is remembered when computing the reference and Δt denotes the sampling time. Note that Equation (1.2) does not take the volatility into account.

The standardized price has now a simple interpretation: when a doubling of the price occurs, the standardized price is +100%. The slope of the price sensitivity line in Figure 1.5a can therefore be expressed in degrees Celsius per percentage of increase of price. This method however assumes that the reference price is different from zero, which is expected to be fulfilled (this can only occur if the price is kept at zero during the whole time window over which the reference is computed).

Combining comfort and flexibility by introducing occupancy modes

Heating needs are different during the day. Having appliances able to control temperature setpoints, it is possible to save energy when users are not home, and to pre-heat in anticipation of their return. Inspired from the Olympic Peninsula experiment (described in Chapter 2), so-called *occupancy* modes are defined, triggered by the user or according to a schedule. Three occupancy modes are introduced: night mode, work mode and home mode. Each occupancy mode is mapped to a setpoint, a price sensitivity and related comfort bounds. For example, during work mode, the heating setpoint could be lowered, and the price sensitivities increased. On the other hand, during night mode, the heating setpoint could stay the same, but the price sensitivity and comfort bounds could be increased to allow for greater flexibility (thus saving money and energy).

1.2.2 Integration into the Danish electricity system

In order to take full benefits of price responsive users, the consumption has to be measured close to real time. The resulting prices adjusting the consumption must then accordingly be sent out right after the measurements have taken place. This puts quite heavy constraints on the IT and communication infrastructure. Metering is right now done by the grid companies, and is far from being realtime. Launching in 2012, Energinet.dk's *DataHub*[2], aims at improving all communication on the Danish electricity market by centralizing the exchanged data and standardizing the data models. However, this might not solve the costly problem of installing the infrastructure needed for a real-time communication with end users.

The control by price concept previously presented therefore aimed at operating on an *aggregated* scale. The individual consumption of each household is not needed: grid measurements (accounting for an aggregate of users) are sufficient because as the number of measured end users grows, their usage patterns become easier to model. It might be necessary to cluster different groups of users in order to see those patterns better (geographically, by house sizes, etc.). Working on aggregates is more robust to missing measurements, due to temporary breakdowns or communication failures. Furthermore, electricity prices can be broadcasted efficiently through the Internet, near real-time. Security measures have to be taken into account, as the authenticity of the price sender has to be verified. This leaves smaller infrastructure costs than extending the electricty grid by the use of

[2] `http://www.energinet.dk/DA/El/DataHub/Sider/DataHub.aspx`

more stressable hardware, reduced to the installation of internet-connected price responsive appliances and to the installation or upgrade of grid measurements instruments that should be able to communicate near real-time. In this project, we will assume near real-time to be 5-minutes. However, this time scale is still an open question and the success of such an implementation will probably directly depend on it.

Wind power largely causes imbalances in the Danish power system, accounting for more than 40% of up- and downregulation imbalances. Because wind power represents such a big portion of the production capacity, its fluctuations have very costly consequences and limitations. As an example, negative prices have been observed on Nord Pool as an effect of a very high wind production at a time where neighbouring countries also have a high wind production. Turbines are then shut down, as they are too expensive to operate compared to the income they generate. In order to be able to integrate even more wind power, the Danish power system has to be better at accommodating with its variability.

A simple but very concrete application of a control by price is therefore to help the Balance Responsible Parties (BRP) reducing penalties they have to pay when deviating from planned production/consumption schedules. If a BRP is able to set up the appropriate metering capabilities of his customers, he could use a price incentive to activate their flexibility. By reducing the paid penalties on the regulating power market, one could say that the BRP sold flexibility to the market.

Breakdown of the electricity price

Seen from the perspective of an end user buying from an electricity supplier, the electricity price consists of several elements. The fraction of electricity price that represents the energy content itself is only approximately 20% of the price that the end user will pay [1], see Figure 1.6. Government taxes represent 37%, Value Added Tax (VAT) 15% and transportation costs (15%).

For an end user to be really price responsive, the variable part of the price must be significant. A political issue is then to change the electricity price structure such that end users can become price responsive. Taxes could for example vary with the Nord Pool prices instead of being fixed.

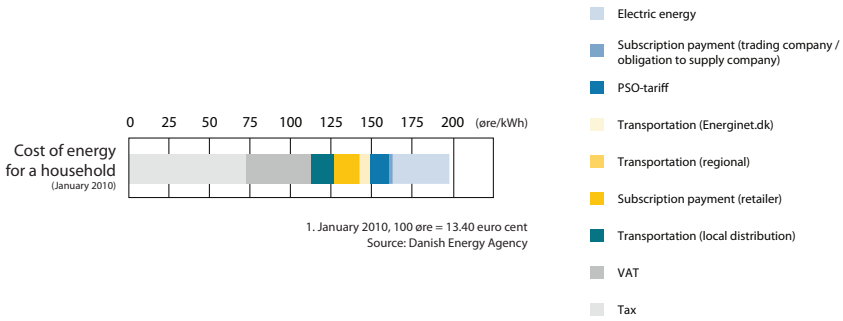


Figure 1.6: *The fraction of the electricity price that represents the energy content itself is only approximately 20%. In order to vary the electricity price, the taxes must also vary, as they account for more than half of the price.*

1.2.3 Related projects

Several projects are related to this topic. The following list is not exhaustive, but is representative of the projects we have been in contact with during this thesis. Project descriptions have directly been taken from the project's websites or introductory reports.

iPower The iPower platform[3] is also known as the Strategic Platform for Innovation and Research within Intelligent Electricity (SPIR). The goal of this platform is to contribute to the development of an intelligent and flexible electricity system capable of handling a large part of sustainable electricity production in areas where production varies due to weather conditions (sun, wind etc). Over the next 5 years, 31 partners shall conduct research and supervise the development of a transition towards production based, flexible electricity consumption as opposed to the current electricity consumption which is based on consumers' needs. The total budget amounts to DKK 120 mio. Risø National Laboratory for Sustainable Energy at the Technical University of Denmark (Risø DTU) is the coordinator and has the overall responsibility that the iPower platform achieves the desired results; whereas Center for Electric Technology (CET) and Danish Technological Institute (DTI) are responsible for coordinating the scientific part and innovation part respectively.

Ensymora The Ensymora (ENergy SYstems MOdelling, Research and Analysis[4])

[3] `http://www.risoe.dtu.dk/Research/sustainable_energy/energy_systems/projects/IES_IPower.aspx`
[4] `http://www.ensymora.dk/`

research project has the objective to develop and improve methods and models used for energy systems analysis and planning. Those models aim at better reflecting the large changes in future energy systems and at analysing technical options, economic incentives, and policies related to both demand and supply of electricity. In particular, the models are used to address the challenges of a fossil free energy system. The project is hosted at the Systems Analysis Division of Risø DTU.

FlexPower FlexPower investigates the potential for using demand as a stable and low cost resource for regulating power. In this project, a market is designed and tested, make using of one-way price signals to activate electricity demand and small-scale generation as regulating power. The idea of using price signals to activate new resources as ancillary services is also be studied. The project runs from June 2010 to June 2013 with a total budget of DKK 10 mio.

EcoGrid The key objective of the EcoGrid EU project[5] is to demonstrate the efficient operation of a distribution power system with high penetration of many and variable renewable energy resources The demonstration will take place on the Danish island Bornholm with more than 50 % electricity consumption from renewable energy production. A real-time market concept will be developed to give small end-users of electricity and distributed renewable energy sources new options (and potential economic benefits) for offering TSO's additional balancing and ancillary services. The total budget for EcoGrid EU is EUR 21 million of which approximately half is financed by the EU. The project is expected to have its formal outset medio 2011.

eFlex The DONG Energy E-Flex project[6] is a demand-response project designed to evaluate the readiness and motivation of household customers to use energy in a flexible manner. A total of 155 qualified households have been identified, and each household is provided with a home energy management solution, which enables the residents to monitor and control the energy consumption, and to individually notice and register the impact and effects which their changes in behaviour can cause in the electricity distribution system and on the environment.

[5] http://energinet.dk/EN/FORSKNING/EcoGrid-EU/Sider/EU-EcoGrid-net.aspx
[6] http://www.dongenergy.dk/distribution/da/privat/eflex/Pages/projekteflex.aspx

1.3 Problem formulation

Consider the following notations:

c_i = Consumption at time index i
p_i = Price at time index i
$\mathbf{z}_i$ = Forecasted external variables at time index i

t = Time index at which a price schedule has to be generated (scheduling time)
k = Lookahead time
K = Maximum forecasting horizon

$\mathbf{c}_t$ = Vector of consumptions up to time t
$\mathbf{p}_t$ = Vector of prices up to time t
$\hat{\mathbf{Z}}_t$ = Matrix of forecasted external variables up to time t
$\mathcal{F}_t$ = Information set available at time t

$p_{t,k}$ = k-step ahead price generated at time t
$c_{t,k}(\cdot)$ = k-step ahead consumption depending on prices generated at time t
$c^*_{t,k}$ = k-step ahead desired consumption target at time t
$w_{t,k}$ = k-step ahead consumption cost weight factor at time t
$p^{\min}_{t,k}$ = k-step ahead lower price limit at time t
$p^{\max}_{t,k}$ = k-step ahead upper price limit at time t
$p^*_{t,k}$ = k-step ahead desired price level at time t
$\lambda_{t,k}$ = k-step ahead penalty associated with deviation from that price level

Every time a control action is to be taken, prices are to be generated up to a horizon K with the objective of following a set of future consumption targets. This can be seen as an optimization problem where future expected costs are minimized, given the information set $\mathcal{F}_t = (\mathbf{c}_t, \mathbf{p}_t, \hat{\mathbf{Z}}_{t+K})$ available at time index t. It is therefore assumed that the external variables have been forecasted up to horizon K.

Because the cost of deviating from the consumption reference varies with time, future costs are weighted with a factor $w_{t,k}$. Furthermore, in order to keep the

price constrained, one could penalize the control signal (imposing a soft constraint) or even restrict it to a certain range. Both solutions are considered here.

Note that a k-step ahead variable associated with a scheduling time t is not uniquely identified by the subscript $t+k$ as several overlapping schedules can refer to the same time index $t+k$. In order to distinguish future values at the same time index $t+k$ associated with different schedules, the subscript t,k is used, denoting the k-step ahead value for the scheduling time t.

At a given time index t, the optimal price schedule $p_{t,1},\ldots,p_{t,k}$ up to a horizon K is obtained by minimizing the (future) expected costs

$$\begin{aligned} L(p_{t,1},\ldots,p_{t,K}) &= \mathbb{E}\Bigg\{ \sum_{k=1}^{K} w_{t,k} \left\| c_{t,k}\left(p_{t,1},\ldots,p_{t,k},\mathcal{F}_t\right) - c^*_{t,k} \right\|^2 \\ &+ \lambda_{t,k} \left\| p_{t,k} - p^*_{t,k} \right\|^2 \Bigg| \mathcal{F}_t \Bigg\} \end{aligned} \tag{1.4}$$

subject to constraints on future prices $p_{t,k}$ (the control signal), and on the future consumption $c_{t,k}$ (the controlled signal)

$$p^{\min}_{t,k} \leq p_{t,k} \leq p^{\max}_{t,k} \tag{1.5}$$

$$0 \leq c_{t,k} < c^{\max}_{t,k}. \tag{1.6}$$

Note that the expectation operator is used in Equation (1.4) because the future consumption $c_{t,k}$ is unknown.

Furthermore, using the fact that $\mathbb{E}\{\cdot^2\} = \mathbb{E}\{\cdot\}^2 + \mathrm{Var}\{\cdot\}$ (from the definition of variance) and because the expectation is a linear operator, one can rewrite the loss function to

$$\begin{aligned} L(p_{t,1},\ldots,p_{t,K}) &= \sum_{k=1}^{K} \Bigg(w_{t,k} \Big(\mathbb{E}\left\{ c_{t,k}\left(p_{t,1},\ldots,p_{t,k},\mathcal{F}_t\right) | \mathcal{F}_t \right\} - c^*_{t,k} \Big)^2 \\ &\quad + w_{t,k} \Big(\mathrm{Var}\left\{ c_{t,k}\left(p_{t,1},\ldots,p_{t,k},\mathcal{F}_t\right) | \mathcal{F}_t \right\} \Big) \\ &\quad + \lambda_{t,k} \left(p_{t,k} - p^*_{t,k} \right)^2 \Bigg). \end{aligned} \tag{1.7}$$

As we will see in Section 5.2.1, the conditional expectation in Equation (1.7) is actually the optimal k-step prediction of the consumption, and the conditional variance is the associated uncertainty.

Minimizing this loss function therefore not only implies finding prices such that the predicted consumption follows the reference: it also implies finding a forecasting model for the consumption, and its related uncertainty. This uncertainty might vary depending on the price, and therefore it must be taken into account when selecting the optimal price schedule.

Four steps have consequently been identified in the development of the project: First, collect data from experiments and simulations. Second, identify the dynamics of the aggregated consumption as a response to a price signal. Third, construct a forecasting model to predict the aggregated consumption. Finally, build a price generator (controller) able to follow a certain consumption target.

CHAPTER 2

The Olympic Peninsula experiment

Up to now, electricity consumers have been primarily 'passive', with predictable and regular consumption patterns. With the necessity of activating consumer flexibility, consumers will play an increasingly important role in the power system. In order to understand and assess the lurking flexibility potentials, field experiments must be conducted. One of those experiments is the *Olympic Peninsula Project*, part of the Pacific Northwest GridWise™ Testbed Demonstration Projects, led as a field demonstration of smart grid technologies by the Pacific Northwest National Laboratory (PNNL) for the U.S. Department of Energy (DOE) [9].

2.1 The experiment

The Olympic Peninsula experiment took place in the Washington and Oregon states between April 2006 and March 2007 aiming at demonstrating to what extent it is possible to control the electricity consumption using electricity prices. The concrete objective was to to constrain the power delivered by the feeder supplying this area. For this purpose, an electricity price signal was sent out every five minutes to residential and commercial energy management systems, inducing actions on the devices under their control. In the light of the project, we want to restrict our focus on residential control. 27 households[1] participated in this project. Firstly, the appliances triggered a control action increasing or lowering the consumption of connected devices depending on the price level. Secondly, those appliances formulated price offers (bids) expressing the current electricity needs of the individual residents. The aggregated electricity bids (demand) and production capacities of the generators (supply) together with feeder constraints (supply limits) yielded a clearance price that was sent out to customers every 5 minute (Figure 2.1). The calculation of the clearing price was done by intersecting supply and demand curves. This means that in this experiment, the *controller* generating the prices was the market itself. To be able to deal with real-cash incentives without changing the existing electricity market structure, a 'shadow' market was established, restricted to the project participants.

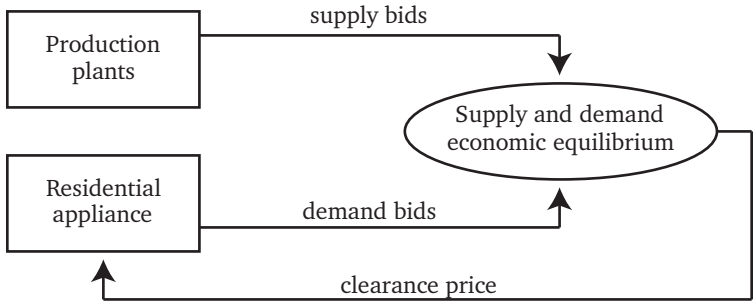


Figure 2.1: *Market control in the Olympic Peninsula experiment. Energy producers and consumers bid on the market according to their needs or available capacity. The resulting clearing price is obtained by matching supply and demand curves leading to an economic equilibrium.*

HVAC (Heating, Ventilation and Air Conditioning) devices, water heaters and cloth dryers of participating customers were controlled in the experiment. The

[1]The Olympic Peninsula experiment report [9] mentions 31 but only 27 are measured

major part of the devices consisted however of heating since not every household was equipped with controlled cloth dryers and space-cooling systems.

Appliances adjusted the consumption of the connected devices according to the concept introduced in Section 1.2, i. e. comfort definitions and price sensitivities of the appliances varied over the day by means of occupancy modes, triggered by the user or according to a schedule. For heating and cooling devices, the interpreted (standardized) price signal caused a shift in the setpoint, depending on a price sensitivity parameter, and associated comfort bounds. For each heating and cooling appliance, each occupancy mode was mapped to a specific setpoint, price sensitivity and their associated comfort bounds. Those mappings were individually assigned by each of the participants before the start of the experiment.

Two groups were available: the Real–Time Pricing (RTP) group of 27 participants described above and a comparative control (CTRL) group of 26 participants which had no price incentives.

The following section describes the available data set in more detail and how it was processed in order to obtain a solid basis for later modelling.

2.2 Preparing the dataset

The following measurements are available, recorded over an experimental period of one year:

- Aggregated energy consumption of the RTP and the comparative CTRL group (15 minutes scale)
- Broadcasted clearance prices (5 minutes scale)
- Occupancy modes of customers (5 minutes scale)
- Initial configurations of comfort parameters depending on occupancy modes
- Weather (temperature, humidity, dewpoint, wind velocity, wind direction, barometer) from three regions (60 minutes scale)
- Solar irradiance data obtained from SolarAnywhere.com [24]

The data suffers from the fact that the time series were partly incomplete, either due to general coarser recording intervals, or to intermittent data losses during recording. Figure 2.2 shows where the missing values are located.

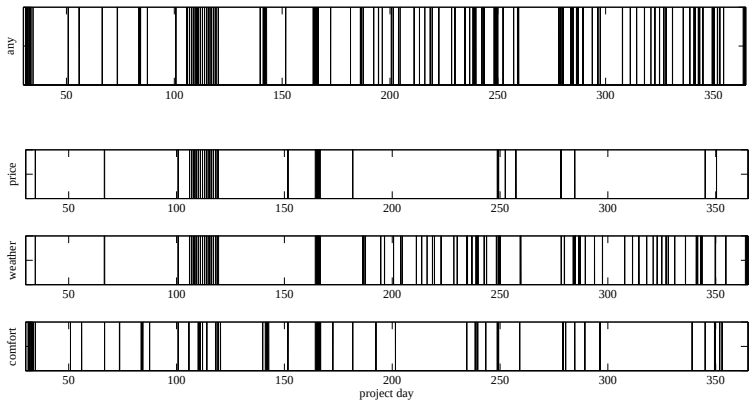


Figure 2.2: *Missing data points are indicated in black area. The top row is showing the times where at least one gap in one of the time series is present, where the missing values for the individual time series are shown in the bottommost three.*

2.2.1 Aligning time scales and handling missing values

As prices and consumption are recorded on different time scales, two datasets are prepared, with two different reference sampling rates: One on a 5-minutes scale and one on a 15-minutes scales. Variables sampled at a lower sampling rate than the reference are interpolated using cubic splines, and variables sampled at a higher sampling rate are averaged out.

For rather slowly varying variables as e. g. the outside temperature, this interpolation seems reasonable and is assumed not to introduce any errors in the time series as long as the interpolation is restricted to a small number of missing samples. But for variables with highly varying dynamics, e. g. the consumption, this interpolation might erroneously introduce unwanted dynamics that lead to difficulties in the modelling stage later on.

Therefore, we restrict our investigation to periods of time where the data set is almost complete, e. g. only a few individual samples of slowly varying variables, such as the weather, are missing. Sufficiently represented periods for the seasons

summer, autumn and winter are then selected (Table 2.1).

Period	Name	# days	# samples (5/15 min scale)	Months considered
1	Summer	19	5472/1824	July, August
2	Autumn	32	9216/3072	October, November
3	Winter	42	12096/4032	January, February, March

Table 2.1: *Selected data periods, where few missing values occur.*

2.2.2 Weather variables

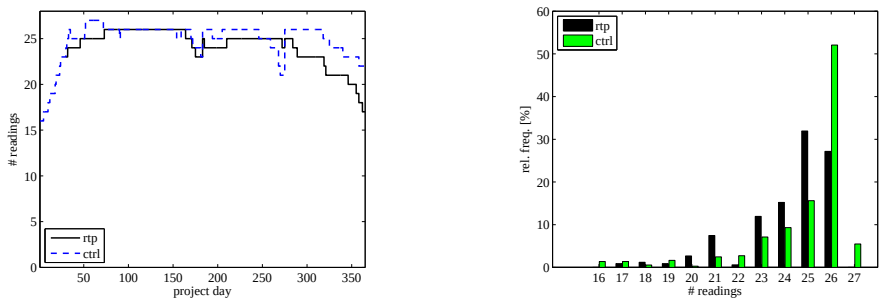
The weather variables are given for three different project regions. However, no corresponding information about the location of individual households was provided. Therefore, all variables related to the weather are averaged over the three project regions in order to obtain a representative variable for later data analysis.

The solar irradiance, however, was not included in the data set. We obtained representative measurements taken at Eugene, Oregon, through SolarAnywhere.com [24].

2.2.3 Estimating the aggregated consumption

At each measurement time, the mean and variance of the individual consumption of each house is recorded. But although the time series of the metered consumption is complete, not all houses where taken into consideration during the measurements (Figure 2.3). This could come from network errors or from not having all the metering devices running at all times.

The aggregated consumption can then be estimated by extrapolating the mean individual consumption alongside its uncertainty (the variance). One should note that the uncertainty (variance) associated with the aggregated consumption evolves proportionally with the square of the number of missing houses.



(a) **(b)**

Figure 2.3: *Even though the time series is complete, the number of houses used in the computation of the mean and variance of the individual consumption of each house varies with time.*

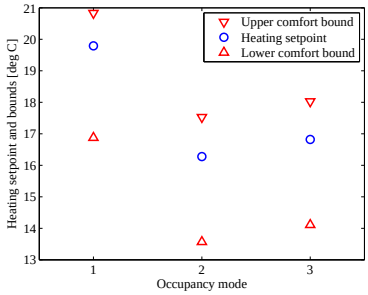
2.2.4 Comfort settings

Customer behaviour was recorded over the entire period of the project by means of a distribution of activated occupancy modes over time, represented by its mean and variance. As mentioned earlier, the participating customers configured cooling and heating devices by assigning those configurations to occupancy modes. For each device, those configurations, or *comfort settings*, consisted for each occupancy mode of the setpoint T_s, the maximum and minimum tolerated setpoint $\{T_s^{\min}, T_s^{\max}\}$, and the parameter k reflecting the price responsivity ($k = 0$ is unresponsive). Since we do not have access to the exact values k_i, we assume a ranking of the indices from lowest to highest responsivity such that

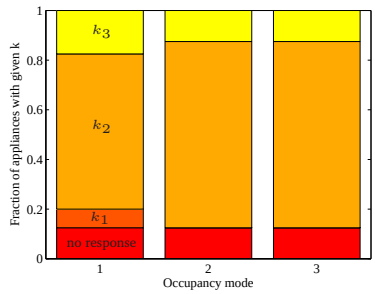
$$0 < k_1 < k_2 < k_3 < \dots \tag{2.1}$$

The switching of comfort settings over time is triggered by 8 occupancy modes $\mathcal{O} \in \{1, \dots 8\}$. However, the association between comfort settings and occupancy modes was fixed once at the project start and was not changed during the project: users did not adjust their comfort settings during time. The mean of those initial assignments is depicted in Figure 2.4. Among the 27 price responsive households, a total number of 40 different devices were initially configured.

Because the three first occupancy modes are triggered 94% of the time, we will focus our analysis on those three, and ignore the others. Those densities, separated into weekdays and weekends, are highly dependent on the time of the day.



(a) *Mean heating set points and mean upper and lower bounds for different occupancy modes.*



(b) *Mean price sensitivities for different occupancy modes.*

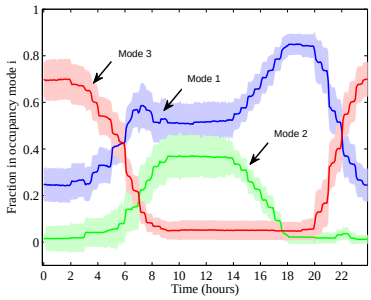
Figure 2.4: *Mean heating setpoints, mean price sensitivities and associated mean comfort bounds for the first three occupancy modes. Averaged over 40 devices. It can be distinguished between a night, home and work mode.*

We distinguish between a night mode, a home mode and a work mode, whereas the latter one is understandably less probable during weekends (Figure 2.5). The number $N_i(t)$ of devices active in a certain occupancy mode $\mathcal{O} = i$ at time t is recorded on a 5 minutes time scale. This allows for estimating a time dependent probability density function $P(\mathcal{O}, t)$ by means of a histogram approximation, i. e. the i'th bin represents the amount of devices activated with mode i. The histogram approximation to the probability density function can be written as

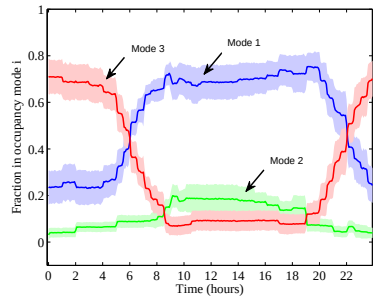
$$P\left(\mathcal{O} = i, t\right) = \frac{N_i(t)}{N_{\text{all}}(t)}, \tag{2.2}$$

where $N_{\text{all}}(t)$ denotes the total number of recorded devices at time t.

Similarly, temperature setpoints and their associated bounds and price sensitivity vary over time. Let us consider the conditional probability density functions $P(T_s|\mathcal{O})$, $P(T_s^{\max}|\mathcal{O})$, $P(T_s^{\min}|\mathcal{O})$ and $P(k|\mathcal{O})$ linking the triggered occupancy mode and the resulting effective parameter. Using the law of total probability, time varying probability density functions describing the evolution of the mean and variance of setpoints, bounds and price sensitivities can be obtained (Equation (2.3)).

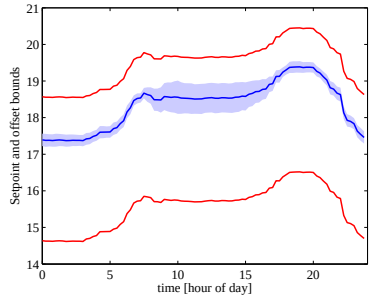


(a) *Weekday*

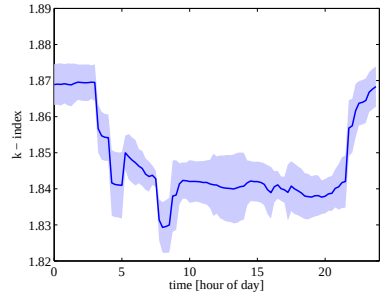


(b) *Weekend*

Figure 2.5: *Three occupancy modes are mainly activated: home mode (1), work mode (2) and night mode (3). The corresponding probabilities of activations are displayed with an area corresponding to one standard deviation.*



(a) *Heating setpoint and offset bounds across one day.*



(b) *Price sensitivities across one day.*

Figure 2.6: *Effective comfort values over one day. Comfort bounds and price sensitivities are rather constant relative to the setpoint.*

$$P(T_s)(t) = \sum_{i=0}^{2} P(\mathcal{O}=i,t)\ \mathcal{D}(T_s|\mathcal{O}=i) \tag{2.3a}$$

$$P(T_s^{\max})(t) = \sum_{i=0}^{2} P(\mathcal{O}=i,t)\ P(T_s^{\max}|\mathcal{O}=i) \tag{2.3b}$$

$$P(T_s^{\min})(t) = \sum_{i=0}^{2} P(\mathcal{O}=i,t)\ P(T_s^{\min}|\mathcal{O}=i) \tag{2.3c}$$

$$P(k)(t) = \sum_{i=0}^{2} P(\mathcal{O}=i,t)\ P(k|\mathcal{O}=i) \tag{2.3d}$$

2.3 Outcome of the experiment

The experiment demonstrated that it is possible to reduce peak-load demand with a simple market-based control. The reduced peak-load is shifted mainly to night hours. The mean consumptions across one day of the RTP and the comparative CTRL group are depicted in Figure 2.7, where the average is taken over the winter period (see Table 2.1). During that period, an average of 5.0% of consumption was shifted to off-peak periods. However, in the process of shifting, consumption was increased by 3.1 pp (percentage points).

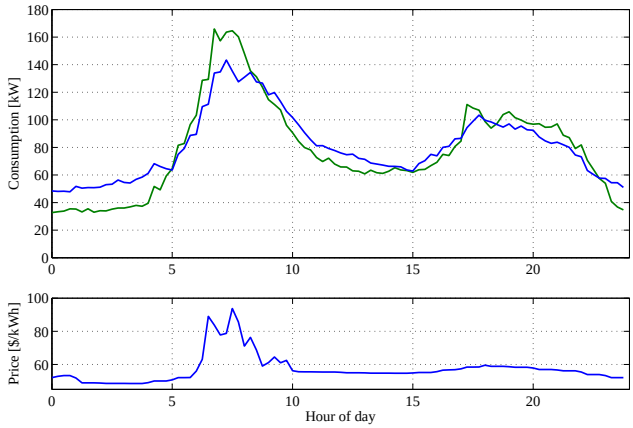


Figure 2.7: *The mean daily pattern of each group shows a successful shift of consumption in order to constrain the feeder. The data used covers 42 days of winter (period 3 in Table 2.1).*

A table summarizing all variables available from this experiment is provided in Appendix A.

CHAPTER 3

Simulating price-responsive consumption

Setting up an experiment is costly requires time and is limited in capabilities. How does the experiment scale with the number of participants? What is the optimal metering resolution needed? How does a „smart" appliance react in a price-sensitive environment?

We lack the tools to answer those questions. Therefore, a *simulation framework* is needed. The aim is not to come up with state-of-the-art appliance and building models: a reasonable and realistic behaviour of the Danish household consumption is sought, as appliance manufacturers and specialized researchers are the best placed to contribute with their knowledge.

3.1 Simulation framework

3.1.1 A modular design

The core of this framework is a Runge-Kutta based Ordinary Differential Equation (ODE) solver, where each building and appliance is described through a set of differential equations. It was designed with modularity in mind, such that each appliance or building model can be tested individually and then easily integrated into the framework. A module consists of a set of states, their associated dynamics (the system of differential equations), events (criteria interrupting the simulation), decisions (state changes when a simulation has been interrupted) and a human behaviour part. For every module, several files are then necessary:

- `<modulename>_Init.m`: contains the initial conditions
- `<modulename>.m`: contains the dynamics
- `<modulename_Event>.m`: evaluates if an event occurred
- `<modulename_HandleEvent>.m`: changes the state of this module according to an event
- `<modulename_GenPtrn>.m`: generates a usage pattern
- `<modulename_EvalConsumption>.m`: based on the state of the module between two samples, computes the mean and integrated consumption between those two samples

In the following discussions, we will distinguish between *samples*, which are times at which the system is measured, and *steps*, which are times at which the system is evaluated or simulated. Sampling times are typically ranging from 5 minutes to 1 hour (we will mostly use 5 minutes), whereas steps are much smaller, typically from couple of seconds to fractions of minutes, in order to precisely account for the system's dynamics. A *debug* mode is also implemented, recording the different variables for each step, and not only for each sample.

Initial conditions

The initialization files contain module parameters and initial states (initial conditions). Both can be sampled from a given distribution in order to vary the

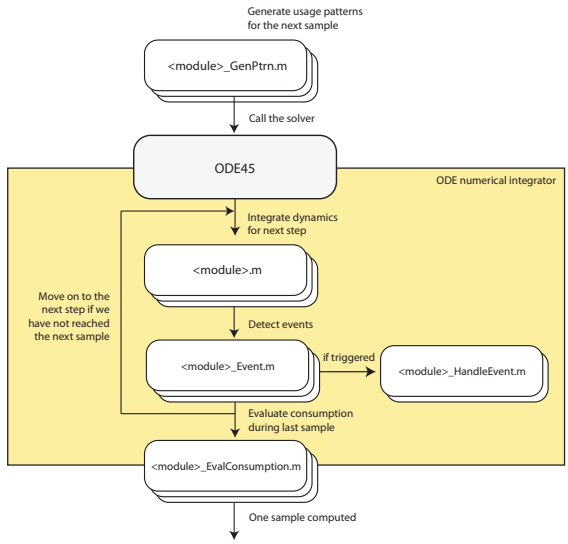


Figure 3.1: *Computation of one sample within the simulation framework. One sample is the result of an iteration over multiple steps of variable length through* MATLAB*'s* `ode45` *solver.*

population of appliances/buildings.

Numerical integration of the dynamics

Let us denote the vector of states of the i'th module $\mathbf{x}^i$, and its associated dynamics $\frac{d}{dt}\mathbf{x}^i = \mathbf{f}^i(\mathbf{x}^i, t)$. By concatenating all the states of all N modules, the set of differential equations describing the whole system can be represented as

$$\frac{d}{dt}\begin{pmatrix}\mathbf{x}^1\\ \vdots \\ \mathbf{x}^N\end{pmatrix} = \begin{pmatrix}\mathbf{f}^1(\mathbf{x}^1, t)\\ \vdots \\ \mathbf{f}^N(\mathbf{x}^N, t)\end{pmatrix} \tag{3.1a}$$

$$\frac{d}{dt}\mathbf{X} = \mathbf{F}\left(\mathbf{X}, t\right) \tag{3.1b}$$

with corresponding initial conditions $\mathbf{X}(t_0) = \mathbf{X}_0 = \left(\mathbf{x}_0^1, \ldots, \mathbf{x}_0^N\right)^T$. It should be noted that the resulting dynamical system is *non-autonomous*, meaning that it also depends on time itself, as external conditions influence the evolution of the system. In order to get the evolution of each state throughout time, a numerical

solver is used. One could use the simple Euler explicit method to express the vector of states at the next step as a linear predictor

$$\mathbf{X}_{n+1} = \mathbf{X}_n + h\mathbf{F}(\mathbf{X}_n, t_n), \tag{3.2}$$

where h is the stepsize. For small h, the error per step is of order h^2 [6]. However, it is often seen that h is chosen to be the sampling interval. This might lead to quite big errors if the states do not evolve linearly in that interval. Furthermore, the stepsize is kept constant. It sounds natural to chose a variable stepsize such that big steps are taken in periods of slow dynamics, since those can reasonably be linearly approximated during a longer time. Considering fast dynamics, a small step size is needed because fast state changes means that the linearization is valid during a shorter time. The `ode45` function in MATLAB, based on a fourth order Runge-Kutta method (also called classical Runge-Kutta), implements a variable step version of the solver, with an error per step of order h^5. This Runge-Kutta method is expressed as

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \tag{3.3}$$

and $t_{n+1} = h + t_n$. The k values are estimates of the slope at the beginning, midpoint, and end of the interval, such that

$$k_1 = h\mathbf{F}(\mathbf{X_n}, t_n) \tag{3.4a}$$

$$k_2 = h\mathbf{F}(\mathbf{X_n} + \frac{1}{2}k_1, t_n + \frac{1}{2}h) \tag{3.4b}$$

$$k_3 = h\mathbf{F}(\mathbf{X_n} + \frac{1}{2}k_2, t_n + \frac{1}{2}h) \tag{3.4c}$$

$$k_4 = h\mathbf{F}(\mathbf{X_n} + k_3, t_n + h). \tag{3.4d}$$

Although requiring four function evaluations, the reduction of error is significant. The dynamics of each module i are then described by a function $\mathbf{f}^i(\mathbf{x}^i, t_n)$ implemented in `<modulename>_.m`.

Events

Between two steps, events can trigger a change of the system: water can be withdrawn from a tank by the user or a heatpump can be turned on or off. The detection of such events is done by having a function `<modulename>_Events.m` returning a true/false boolean value whether or not an event has been triggered.

If an event is triggered, the `<modulename>_HandleEvents.m` function is called, altering the state of the module triggering the event. For example, if the temperature in a house is so low that it requires turning on heating, an event is triggered, and the state element controlling whether or not heating is supplied is changed.

Usage pattern

At each sample, a usage pattern is generated for the next sample. This usage pattern can for example be the amount of hot water used in a household during one sample or a change in heating setpoints triggered by a new occupancy mode. The function `<modulename>_GenPtrn.m` is called at each sample, generating this pattern and altering the state of this module accordingly.

3.1.2 Description of modules

The dynamics of the whole system can be broken down into two parts: the heating dynamics of the building, and the dynamics of each appliance inside that building. Those dynamics have coupled variables (e. g. indoor air temperature) and depend on both external variables (e. g. external air temperature) and on a certain customer behaviour model (e. g. hot water usage), as seen in Figure 3.2. It therefore makes sense to create a parent/child relation between appliances and buildings because each appliance can belong to one building only. We will also restrict ourselves to the case where no more than one type of heating appliance is present for each parent building.

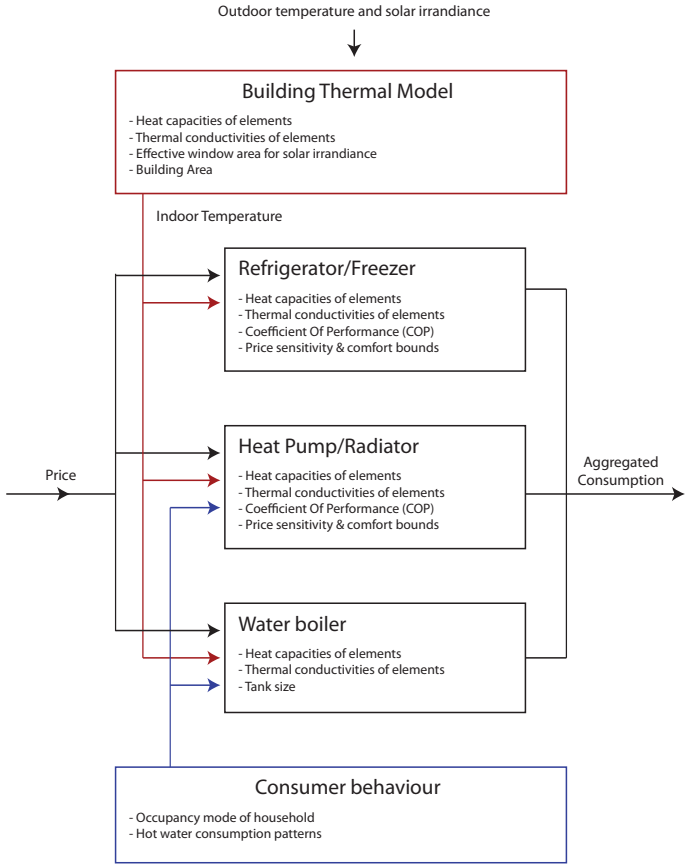


Figure 3.2: *Modular design of the simulation environment*

Short heat transfer theory

As a form of energy, heat has the unit joule (J) in the International System of Units (SI). The standard unit for the rate of heat transferred is the Watt (W), defined as joules per second.

For incompressible substances, heat can be defined as the amount of energy needed to change the temperature of that substance from T_0 to T_f.

$$Q = \int_{T_0}^{T_f} C \, dT = C(T_f - T_0) = C\Delta T, \tag{3.5}$$

where C is the heat capacity (assumed independent of temperature), expressed in joules per degree Celsius (or equivalently in joules per Kelvin). It represents the amount of heat needed to heat up the substance with one degree. It can be convenient to define the heat capacity per unit of mass, per unit of volume, or per unit of area (in the case of a section). Considering heat capacity per unit of mass, the term *specific* heat capacity is employed, represented by the lowercase letter c

$$Q = mc\Delta T. \tag{3.6}$$

Considering an infinitesimal change in temperature dT where the mass m is constant, the change of heat can be defined as

$$dQ = mcdT \;\Leftrightarrow\; \frac{dQ}{dt} = mc\frac{dT}{dt}., \tag{3.7}$$

Let us now consider two elements, where the heat conduction within those is much faster than the heat conduction across the boundary of the elements. This means that a *lumped capacitance approximation* can be used, by reducing one aspect of the transient conduction system (the one within the object) to an equivalent steady state system. In other terms, only the heat conduction between those elements remains, as the heat conduction inside them is approximated by a steady state system, resulting in a temperature within the object being completely uniform, although its value may be changing with time.

The law of heat conduction, also known as Fourier's law, describes the time evolution of the amount of heat exchanged between two elements at temperatures T_1 and T_2 with contact surface area A

$$\frac{dQ}{dt} = -\lambda A\,(T_1 - T_2) = U\,(T_2 - T_1) \tag{3.8}$$

where λ is the heat conductivity, expressed in joules per seconds, per degree Celsius (or Kelvin) and per contact surface area between the two elements. For simplicity, we will denote $U = \lambda A$ as the heat conductivity between element 1 and 2. Note that the heat flow dQ/dt is directed positively from T_2 to T_1.

Taking element 1 as reference, its change of heat from equation (3.7) is equal to the sum of all heat flows applied on it. Having only one heat flow (the one between element 1 and 2), the time evolution of its temperature can be described by combining Equation (3.7) and Equation (3.8):

$$\frac{dQ_1}{dt} = m_1 c_1 \frac{dT_1}{dt} = U\left(T_2 - T_1\right) \tag{3.9}$$

In a system consisting of n elements i with masses m_i, specific heat capacities c_i and all exchanging heat between each others with conductivities U_{ij}, we can write up the time derivative of the temperature of element i as the sum of all heat flows applied to it.

$$m_i c_i \frac{dT_i}{dt} = \sum_{j=1}^{n} \frac{dQ_{ij}}{dt} \tag{3.10}$$

If all heat transfers follow Fourier's law, and by using the heat capacity $C_i = m_i c_i$,

$$C_i \frac{dT_i}{dt} = \sum_{j=1}^{n} U_{ij}\left(T_j - T_i\right). \tag{3.11}$$

Thermal model of a building

Based on [17], [20] and [2], the thermal dynamics of a building can be modelled as two individual thermal masses: one corresponding to the inside air and furniture, and the other corresponding to the structure of the building (Figure 3.3). The building structure is subject to heat exchanges with outside air. The inside air is also subject to heat exchanges with outside air (due to natural ventilation of the house) but is also affected by outside solar irrandiance through the windows.

This yields the following system of differential equations governing the evolution of the temperatures of the building T_b and of the inside air T_a:

$$C_a \frac{dT_a}{dt} = U_{ao}\left(T_o - T_a\right) + U_{as}\left(T_b - T_a\right) + A_w S + P \tag{3.12a}$$

$$C_b \frac{dT_b}{dt} = U_{so}\left(T_o - T_b\right) + U_{as}\left(T_a - T_b\right) \tag{3.12b}$$

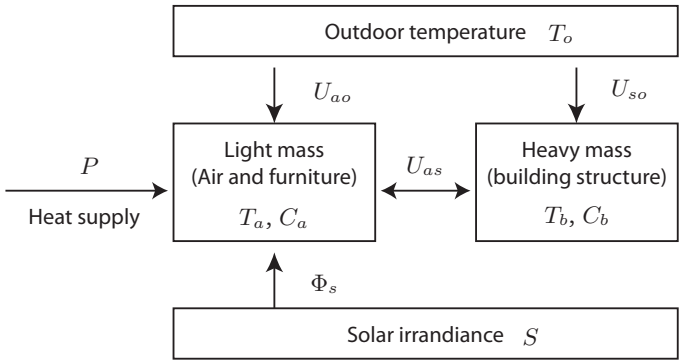


Figure 3.3: *Thermal model of a building*

For a square building of area A and height h, assuming a 20% window surface gives a solar irrandiance surface of $A_w = 0.2 \cdot h\sqrt{A}$. If only one of the four walls is exposed to the sun, the heat flow originating from the irrandiance of the sun is $\Phi_s = 0.05 \cdot hI\sqrt{A}$, which is in accordance with the FlexHouse modelled in [2].

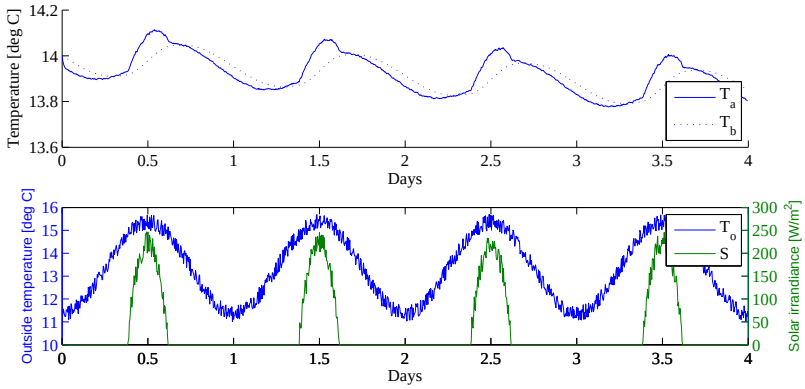


Figure 3.4: *The inside air temperature T_a and building temperature T_b are here exposed to a fluctuating outside temperature T_o and sun irradiance S. No space heating is applied. The transfer function of the building can be characterized as a lowpass filter with a corresponding phase shift (delay).*

Regarding the living area parameter A, its distribution is based on Danish statistics [25]. A continuous density function was obtained by smoothing a histogram of the probabilities associated with groups of area size. A Gaussian kernel was

used, with a manually selected bandwidth (Figure 3.5).

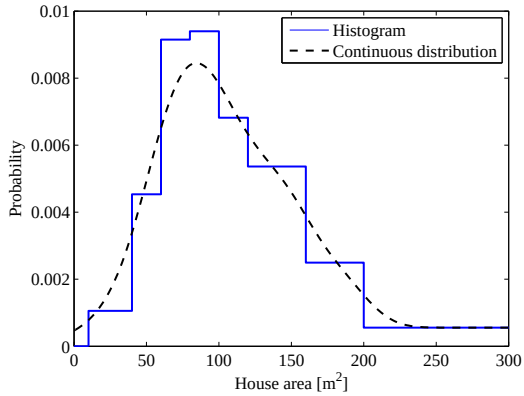


Figure 3.5: *Probability density function of Danish house areas taken from [25].*

Sampling areas from this arbitrary distribution can be done using the *inverse* transform sampling [11]. This derives from the probability integral transform, stating that if a stochastic variable X is transformed by its cumulative distribution function F_X, then the resulting stochastic variable $Y = F_X(X)$ is uniformly distributed on $[0, 1]$. By generating uniformly distributed numbers y on $[0, 1]$, and by taking x as the value for which $F_X(x) = y$, then one is sampling from $F_X = F_X^{-1}(Y)$.

The rest of the parameters are distributed according to [17] and [20].

Symbol	Description	Value
A	Floor area	Sampled from Danish statistics
A_s	Surface area	$2 \cdot A + 4h\sqrt{A}\ \mathrm{m}^2$
A_w	Window area	$0.05 \cdot h\sqrt{A}\ \mathrm{m}^2$
C_a	Air and furniture heat capacity	$13 \cdot A\ \mathrm{kJ/^\circ C}$
C_s	Building structure heat capacity	$\mathcal{N}(360, 70) \cdot A\ \mathrm{J/^\circ C}$
U_{ao}	Heat transfer coefficient between inside and outside air	$0.42 \cdot A \quad \mathrm{J/s/^\circ C}$
U_{so}	Heat transfer coefficient between building structure and outside air	$\frac{7.69 \cdot A_s\,(1.07 \cdot A + 69.0)}{7.69 \cdot A_s - 1.07 \cdot A - 69.0}\ \mathrm{J/s/^\circ C}$
U_{as}	Heat transfer coefficient between inside air and building structure	$7.69 \cdot A_s\ \mathrm{J/s/^\circ C}$

Table 3.1: *Thermal model parameters for the building*

Generating user behaviour in buildings

Heating setpoints are assigned for each occupancy mode, the latter being triggered during the day following a certain pattern (Figure 2.5). We seek to build a model reproducing this behaviour to represent the varying heating needs during the day. First of all, it is assumed that the assignment of heating setpoints according to each occupancy mode is randomly distributed following

$$P(T_s|\mathcal{O}=i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2} \tag{3.13}$$

where the parameters μ_i ans σ_i are estimated from the Olympic Peninsula data set. The activation of the occupancy modes is slightly more difficult. At all times t of the day (represented by the minute of the day), there is a probability $x_{t,i}$ of being in a mode i. The transition from a set of probabilities $\mathbf{x_t} = (x_{t,1}, x_{t,2}, x_{t,3})^T$ to the next sample $\mathbf{x_{t+1}}$ can be written as a Markov Chain with a finite state space:

$$\mathbf{x}_{t+1|t}^T = \mathbf{x}_t^T \mathbf{A}_t \tag{3.14}$$

with the two conditions

$$\sum_{i=1}^{3} \text{row}_i(\mathbf{A}_t) = 1 \tag{3.15}$$

$$\forall i = 1,\dots,3,\ j = 1,\dots,3,\ 0 \leq a_{t,ij} \leq 1 \tag{3.16}$$

Note that the transition matrix $\mathbf{A}_t$ is a 3×3 time dependent matrix. Equation (3.15) describes 3 linear relationship between the 3 variables, meaning that the number of variables to be estimated reduces from 9 to 6. Defining the vector of unknown as $\boldsymbol{\theta_t} = (\theta_{t,1}, \dots, \theta_{t,6})^T$, the linear dependency can be written as

$$\mathbf{A}_t = \begin{pmatrix} \theta_{t,1} & \theta_{t,2} & 1-\theta_{t,1}-\theta_{t,2} \\ \theta_{t,3} & \theta_{t,4} & 1-\theta_{t,3}-\theta_{t,4} \\ \theta_{t,5} & \theta_{t,6} & 1-\theta_{t,5}-\theta_{t,6} \end{pmatrix} \tag{3.17}$$

By defining

$$\mathbf{M}_t = \begin{pmatrix} x_{t,1} & 0 & x_{t,2} & 0 & x_{t,3} & 0 \\ 0 & x_{t,1} & 0 & x_{t,2} & 0 & x_{t,3} \\ -x_{t,1} & -x_{t,1} & -x_{t,2} & -x_{t,2} & -x_{t,3} & -x_{t,3} \end{pmatrix}$$

and

$$\mathbf{b}_t = \begin{pmatrix} 0 \\ 0 \\ x_{t,1}+x_{t,2}+x_{t,3} \end{pmatrix},$$

Equation (3.14) can then be rewritten as

$$\mathbf{x}_{t+1} = \mathbf{M}_t \boldsymbol{\theta}_t + \mathbf{b}_t, \tag{3.18}$$

keeping only the constraint stating that every parameter in $\boldsymbol{\theta}$ should be between 0 and 1. Stacking all the measurements at time t and $t+1$ vertically in $\mathbf{x}_{t+1}$, $\mathbf{M}_t$ and $\mathbf{b}_t$, the parameters $\boldsymbol{\theta}$ can be estimated by using linear least-squares [15]. Note that two sets of matrices $\mathbf{M}_t$ have to be estimated: one for weekends, and one for weekdays, as they show different occupancy patterns (Figure 2.5). Finally, the parameters $\boldsymbol{\theta}$ can be used to reconstruct the matrix $\mathbf{A}_t$.

For the purpose of generating occupancy modes during the day, an initial value at time t is sampled from the distribution of occupancy modes at time t, obtained by averaging over all measurements at the minute of the day t. The initial mode is then converted to a boolean row vector where a 1 is put into the column corresponding to the activated occupancy mode. Activating occupancy mode 3 yields for example the vector $\mathbf{x}_t = (0, 0, 1)$. Given that we are in occupancy mode i, we can calculate the probability distribution of transitioning to mode j by

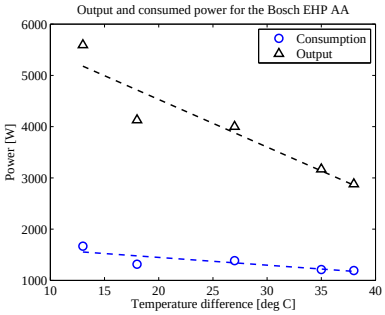
$$P(\mathcal{O}_{t+1} = j | \mathcal{O}_t = i) = a_{t,ij} \tag{3.19}$$

meaning that the next occupancy mode can be obtained by sampling from a distribution being the i'th row of $\mathbf{A}_t$ (because the current occupancy mode is i). Sampling from a specific probability density function can be done by the inverse transform sampling method, described in the building heating model. Note that instead of estimating individually the transitions for one minute of the day to the other, one could describe the matrix $\mathbf{A}_t$ with a model directly depending on the minute of the day, using for example a finite number of Fourier components.

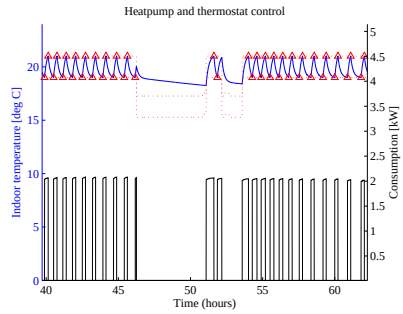
Air to air heat pump model

An air to air heatpump is assumed to deliver an amount P of power directly to the air. One should note that this is a quite heavy approximation, because in real life, power is not delivered uniformly and instantly to the air. The heatpump power is dependent on the difference between outside and inside temperature, and that dependency is linearly approximated [17] [20] for the Bosch EHP AA used here (Figure 3.6a).

The temperature inside the room is then controlled so that is does not deviate more than the the hysterisis, T_{hys}, from a certain setpoint T_s. The heatpump is



(a) *Measured consumption and output for the Bosch EHP AA*



(b) *The indoor air temperature is controlled by turning on and off the heatpump so that the temperature does not deviate more than T_{hys} from the setpoint T_s*

Figure 3.6

therefore activated when the indoor air temperature drops below $T_s - T_{hys}$, and stopped when the indoor air temperature increases over $T_s + T_{hys}$, as seen in Figure 3.6b. Note that the heatpump control system is here modelled as a simple on/off switch. More advanced methods could of course be considered.

Symbol	Description	Value
T_s	Temperature setpoint	dependent on occupancy mode
$T_s^{\max}$	Maximum temperature setpoint	dependent on occupancy mode
$T_s^{\min}$	Minimum temperature setpoint	dependent on occupancy mode
T_{hys}	Temperature setpoint hysteresis	$1\ ^\circ\mathrm{C}$
$P(T_o - T_a)$	Supplied power	linear relation, as in Figure 3.6a
C_{on}	HP's consuption	linear relation, as in Figure 3.6a

Table 3.2: *Parameters for the heat pump*

Fridge model

One possibility to model a fridge is to consider it as a system consisting of three parts with different thermal masses [20]: a very light mass (the cooling circuit), a light mass (the fridge content and interior air) and a heavy mass (the fridge itself with its insulating structure).

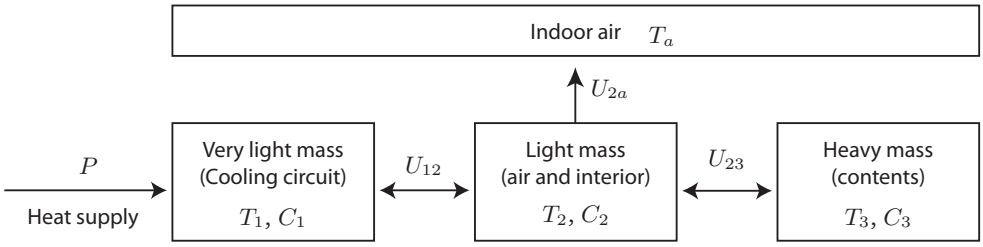


Figure 3.7: *Thermal model of a fridge*

The set of equations describing the dynamics of such a system is as following

$$C_1 \frac{dT_1}{dt} = U_{12}(T_2 - T_1) + P \tag{3.20}$$

$$C_2 \frac{dT_2}{dt} = U_{12}(T_1 - T_2) + U_{23}(T_3 - T_2) + U_{2a}(T_a - T_2) \tag{3.21}$$

$$C_3 \frac{dT_3}{dt} = U_{23}(T_2 - T_3) \tag{3.22}$$

$$\tag{3.23}$$

The supplied heat flow P is defined as $-\mathrm{COP} \cdot C_{\mathrm{on}}$ where COP is the coefficient of performance, and C_{on} is the consumption of the fridge when turned on. The coefficient of performance is defined as the ratio between electrical power consumed and amount of energy provided by the fridge compressor.

Several parameters are distributed to represent different fridge sizes, loadings and efficiencies (Table 3.3).

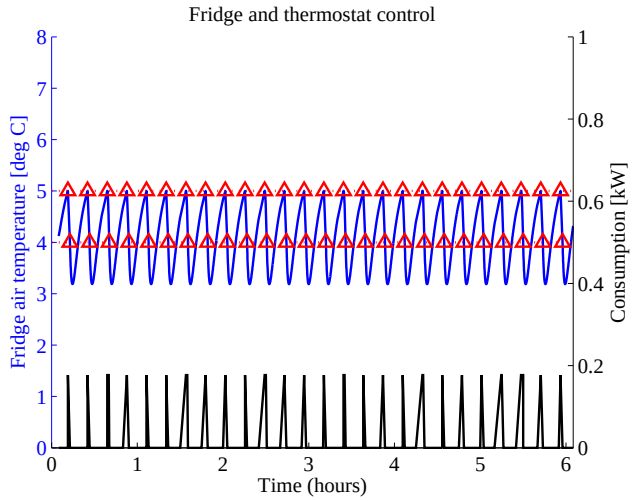


Figure 3.8: *The inside air temperature of the fridge is controlled by turning on and off the compressor so that the temperature does not deviate more than T_{hys} from the setpoint T_s.*

Symbol	Description	Value
T_s	Temperature setpoint	$4.5\ ^\circ\mathrm{C}$
T_{hys}	Temperature setpoint hysteresis	$0.5\ ^\circ\mathrm{C}$
C_1	Heat capacity of cooling circuit	$1\ \mathrm{kJ}/^\circ\mathrm{C}$
C_2	Heat capacity of fridge air	$\mathcal{N}(13, 1)\ \mathrm{kJ}/^\circ\mathrm{C}$
C_3	Heat capacity of fridge contents	$\mathcal{N}(251, 1750)\ \mathrm{kJ}/^\circ\mathrm{C}$
U_{2a}	Heat transfer coefficient between fridge air and indoor air	$\mathcal{N}(5, 0.25)\ \mathrm{J/s}/^\circ\mathrm{C}$
U_{12}	Heat transfer coefficient between 1 and 2	$\mathcal{N}(12, 1)\ \mathrm{J/s}/^\circ\mathrm{C}$
U_{23}	Heat transfer coefficient between 2 and 3	$30\ \mathrm{J/s}/^\circ\mathrm{C}$
COP	Coefficient of performance	$\mathcal{N}(2.8, 0.04)$
C_{on}	Consumption when turned on	$\mathcal{N}(200, 2500)\ \mathrm{W}$

Table 3.3: *Parameters for the fridge*

Combined fridge/freezer model

The same approach can be used to model a *combined* fridge and freezer. Using the model proposed in [23], a Hotpoint Iced Diamond RSB20 120L (10L freezer) unit is modelled. The temperature of the freezer is left uncontrolled: only the fridge temperature is monitored and controlled.

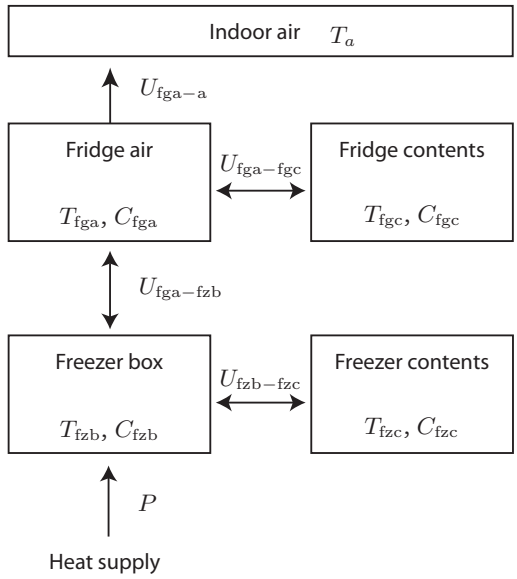


Figure 3.9: *Thermal model of a combined fridge/freezer.*

According to Figure 3.9 we can set up the following system of equations

$$C_{\mathrm{fga}}\frac{dT_{\mathrm{fga}}}{dt} = U_{\mathrm{fga-a}}\left(T_a - T_{\mathrm{fga}}\right) + U_{\mathrm{fga-fgc}}\left(T_{\mathrm{fgc}} - T_{\mathrm{fga}}\right) + U_{\mathrm{fga-fzb}}\left(T_{\mathrm{fzb}} - T_{\mathrm{fga}}\right) \quad (3.24)$$

$$C_{\mathrm{fgc}}\frac{dT_{\mathrm{fgc}}}{dt} = U_{\mathrm{fga-fgc}}\left(T_{\mathrm{fga}} - T_{\mathrm{fgc}}\right) \quad (3.25)$$

$$C_{\mathrm{fzb}}\frac{dT_{\mathrm{fzb}}}{dt} = U_{\mathrm{fga-fzb}}\left(T_{\mathrm{fga}} - T_{\mathrm{fzb}}\right) + U_{\mathrm{fzb-fzc}}\left(T_{\mathrm{fzc}} - T_{\mathrm{fzb}}\right) + P \quad (3.26)$$

$$C_{\mathrm{fzc}}\frac{dT_{\mathrm{fzc}}}{dt} = U_{\mathrm{fzb-fzc}}\left(T_{\mathrm{fzb}} - T_{\mathrm{fzc}}\right) \quad (3.27)$$

where T_{fga} denotes the fridge's inside air temperature, T_{fgc} the fridge's content temperature, T_{fzb} the freezer's box temperature and T_{fzc} the freezer's content temperature.

The controlled variable is the fridge air temperature T_{fga}, which should lie within $T_s \pm T_{hys}$. The supplied power P is defined as $-\mathrm{COP} \cdot C_{\mathrm{on}}$ where COP is the coefficient of performance, and C_{on} is the consumption of the combined fridge/freezer when turned on.

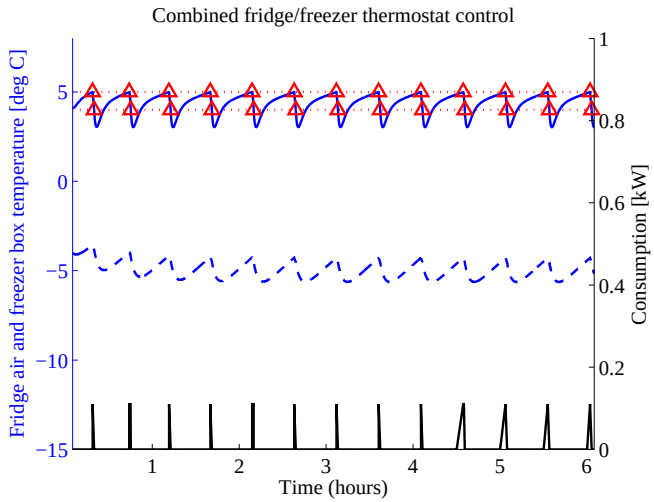


Figure 3.10: *The inside air temperature of the fridge is controlled by turning on and off the compressor so that the temperature does not deviate more than T_{hys} from the setpoint T_s.*

Several parameters are distributed to represent different combined fridge/freezer sizes, loadings and efficiencies (Table 3.4).

Symbol	Description	Value
T_s	Temperature setpoint	$4.5\ ^\circ\mathrm{C}$
T_{hys}	Temperature setpoint hysteresis	$0.5\ ^\circ\mathrm{C}$
C_{fga}	Heat capacity of fridge air	$\mathcal{N}(0.5, 0.0025)\ \mathrm{kJ}/^\circ\mathrm{C}$
C_{fgc}	Heat capacity of fridge contents	$\mathcal{N}(4, 0.01)\ \mathrm{kJ}/^\circ\mathrm{C}$
C_{fzb}	Heat capacity of freezer box	$\mathcal{N}(1.350, 0.04)\ \mathrm{kJ}/^\circ\mathrm{C}$
C_{fzc}	Heat capacity of freezer contents	$\mathcal{N}(6, 1)\ \mathrm{kJ}/^\circ\mathrm{C}$
$U_{\mathrm{fga-a}}$	Heat transfer coefficient between fridge air and indoor air	$\mathcal{N}(1.2, 0.0001)\ \mathrm{J/s}/^\circ\mathrm{C}$
$U_{\mathrm{fga-fgc}}$	Heat transfer coefficient between fridge air and fridge contents	$\mathcal{N}(4.375, 0.01)\ \mathrm{J/s}/^\circ\mathrm{C}$
$U_{\mathrm{fga-fzb}}$	Heat transfer coefficient between fridge air and freezer box	$\mathcal{N}(1.35, 0.0025)\ \mathrm{J/s}/^\circ\mathrm{C}$
$U_{\mathrm{fzb-fzc}}$	Heat transfer coefficient between freezer box and freezer contents	$\mathcal{N}(1.8750, 0.0025)\ \mathrm{J/s}/^\circ\mathrm{C}$
COP	Coefficient of performance	$\mathcal{N}(2.8, 0.2)$
C_{on}	Consumption when turned on	$\mathcal{N}(100, 20)\ \mathrm{W}$

Table 3.4: *Parameters for the combined fridge/freezer*

Electric water boiler model

Based on [10], an electric water heater with a single heating element can be modeled as shown in Figure 3.11.

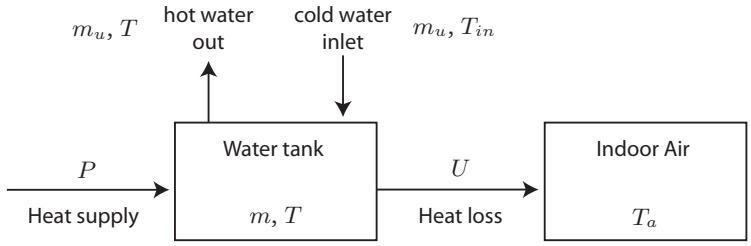


Figure 3.11: *Thermal model of a water heater.*

The water in the tank of mass m and temperature T, is assumed to be heated up by an electrical boiler with power P. Additionally, the tank is subject to heat exchanges with the ambient air temperature, characterized by the insulation of the heater having heat conductivity U.

Assuming no water being withdrawn from the tank, the differential equation describing the evolution of the water temperature T follows

$$mc_w \frac{dT}{dt} = U\left(T_a - T\right) + P \tag{3.28}$$

where c_w is the specific heat capacity of water[1].

It is assumed that when a mass m_u of hot water is used, it is immediately replaced by an equal quantity of cold water. Let us define the mixing temperature T_m as as the average of the hot and cold temperatures, weighted by their associated masses. The amount of heat removed from the inside water can be seen as the

[1] Specific heat capacity of water: 4180 J/kg/C assuming a specific density of water of 0.990 kg/L at 20 °C.

amount of heat removed by cooling down from T to the mixing temperature T_m.

$$\begin{aligned} Q &= \int_T^{T_m} mc_w \, dT && (3.29)\\ &= mc_w \left(T_m - T\right) && (3.30)\\ &= mc_w \left(\frac{\left(m - m_u\right) T + m_u T_{in}}{m} - T\right) && (3.31)\\ &= c_w \left(m - m_u\right) T + m_u c_w T_{in} - mc_w T && (3.32)\\ &= m_u c_w \left(T_{in} - T\right) && (3.33) \end{aligned}$$

Knowing the relation $Q = mc_w \Delta T$, if the amount of heat Q is transferred during an infinitesimal amount of time dt (e. g. the mixing process is instantaneous), then the infinitesimal amount of heat dQ transferred is proportional to the infinitesimal amount of water mass dm_u replaced, such that

$$\begin{aligned} dQ &= dm_u c_w \left(T_{in} - T\right) && (3.34)\\ \frac{dQ}{dt} &= \frac{dm_u}{dt} c_w \left(T_{in} - T\right). && (3.35) \end{aligned}$$

The overall heat flow of the system is a sum of the heat flow through the tank insulation, the heating power from the boiler, and the heat flow through water being withdrawn. Adding those heat flows together, we obtain the equation governing the dynamics of this system:

$$mc_w \frac{dT}{dt} = U \left(T_a - T\right) + P + \frac{dm_u}{dt} c_w \left(T_{in} - T\right). \tag{3.36}$$

We will here assume that the tank size in each house depends on the number of persons living in the household. The joint distribution of number of persons in a household and living area can be obtained at Danmark's Statistik [25]. In order to generate a representative set of the number of persons per house, we sample from this joint distribution, shown in Figure 3.13. This is done by using the inverse transform sampling, previously described in the building heating model.

Knowing the number of persons n_{pers} per house enables a rough approximation of the storage capacity which should be chosen in order to cover rush hours of hot water usage. The tank size is calculated as the amount of hot water needed in peak demand. It is assumed that the peak demand is reached when all house members take a shower after each other (laundry or dishwasher assumed to be off during that period). It has been found that an average shower takes eight

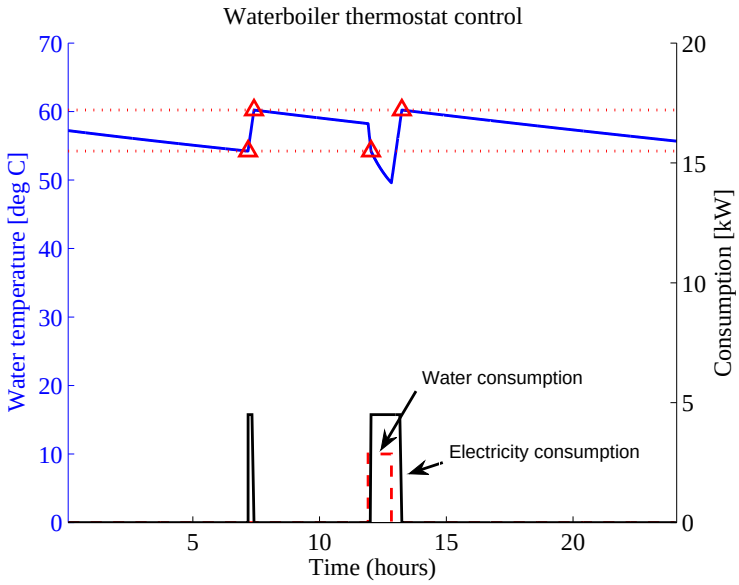


Figure 3.12: *The inside water temperature of the boiler is controlled by turning on and off the heating element so that the temperature does not deviate more than T_{hys} from the setpoint T_s.*

minutes and that the average flow rate is 9.5 L/min [2], so the approximate tank size is found to be

$$\mathrm{Vol}_{\max} = n_{\mathrm{pers}} \times 8 \text{ min} \times 9.5 \text{ L/min}. \tag{3.37}$$

According to this tank volume, the electricity consumption of the heating element can be set. Different models of water heaters have been investigated, yielding consumptions from 3800 to 6000 Watts, depending on the storage capacity. A COP value of approximately one[3] is used, because energy is completely converted to heat. The heat transfer coefficient between the tank and ambient air is obtained from [10].

The following table summarizes the parameters being distributed within the simulations:

[2] International Association of Plumbing and Mechanical Officials (IAPMO), *Uniform Plumbing Code*, 2011, `http://en.wikipedia.org/wiki/Uniform_Plumbing_Code.pdf`

[3] Reliant Energy 2005, `http://www.reliant.com/en_US/Platts/PDF/P_PA_26.pdf`

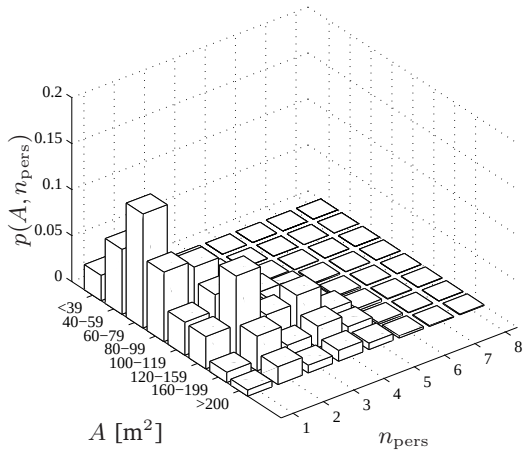


Figure 3.13: *Joint probability density function of house areas A and number of persons n_{pers} in the household, based on Danish statistics [25].*

Symbol	Description	Value
$\mathrm{Vol}_{\max}$	Tank volume	dependent on house area A
T_s	Temperature setpoint	$\mathcal{N}(45, 2.5)$ °C
T_{hys}	Temperature setpoint hysteresis	2.778 °C
T_c	Temperature of inlet water	$\mathcal{N}(10, 2.25)$ °C
COP	Coefficient of performance	1
C_{on}	Consumption when turned on$_{\max}$)	dependent on tank volume $Vol_{\max}$
U	Heat transfer coefficient between tank and ambient air	$\mathcal{N}(1.1, 0.03)$ J/s/°C

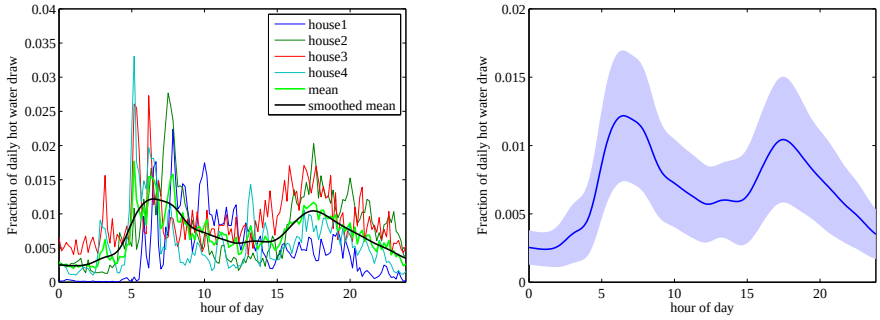
Table 3.5: *Parameters for the water heater simulation.*

Hot water usage patterns

In order to approximate residential hot water usage profiles we used data from four Danish households, which were recorded over a period of one year [5]. The water consumption data consists of average values over intervals of 10 minutes. Note that these hot water usage profiles were found by separating hot water usage from the overall district heating consumption (including hot water used to heat) using statistical methods in [5].

An average of the four houses' profiles is taken, after being normalized by the number of persons in the household. The mean profile is smoothed out using a Gaussian kernel in order to obtain a distribution which we can sample from. The computation is carried out for weekdays and weekends separately (Figure 3.14a). The resulting profile strongly resembles observations from [21].

For each simulation sample, each person is assumed to use hot water according to a time varying Gaussian distribution, characterized by a time–varying mean and its associated time–varying standard deviation. This distribution is expressed in fractions of daily consumption per inhabitant, the latter found to be approximately 50 litres[4] (but this could very well be extended to include dependencies on the age of the inhabitant for example). The resulting distribution for weekday water consumption is depicted in Figure 3.14b.



(a) *Individual and average hot water usage profiles during weekdays.*

(b) *Time varying Gaussian distribution extracted, characterized its mean and variance on a time scale of 5 minutes.*

Figure 3.14

[4] `http://www.umweltbewusst-heizen.de`

3.2 Preparing and validating a dataset

Preparing a dataset requires having access to realistic weather conditions. Furthermore, it requires a realistic description of the population of consumers. Those requirements, and how we obtained the relevant information, are listed below:

Initial configuration of setpoints Taken from the Olympic Peninsula data set. Depends on the occupancy mode.

Price sensitivities Kept constant at -2 °C by unit of relative price change (e. g. a doubling of the price yields a 2 °C decrease).

Comfort bounds Maximum and minimum deviations allowed from the setpoint are set to ± 2 °C.

Weather conditions Outside temperature and sun irrandiance at an hourly resolution was supplied by DMI[5]. The geographical region considered is Copenhagen, and the data set consists of *forecasted* values, initially computed every 6 hours for a horizon of 24 hours.

Price interpretation Every appliance is assumed to interpret the received price in the same way, meaning that they all use the same smoothing constant τ in the computation of the standardized price ρ.

Sampling rate We have assumed that the aggregated consumption can be measured every 5 minutes.

Appliances A first approach is taken by only simulating heating appliances.

Prices Various steps of price of random magnitudes and random durations are generated as inputs.

A dataset consisting of two month of data, starting 1st of January 2009 is then simulated, containing $2 \times 60 \times 24 \times 30/5 = 17280$ data points. No matter what the initial states are, the system converges after a couple of days (Figure 3.15).

The framework is optimized to minimize computing time and to ensure scalability in the number of houses/appliances. Running times are satisfactory (Figure 3.16a) as it seems to run in polynomial time.

[5]Danish Meteorological Institute

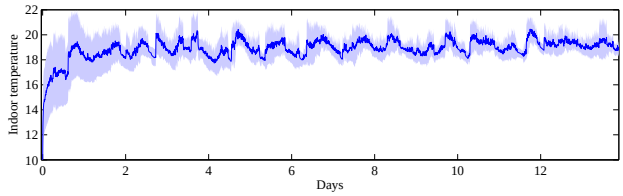
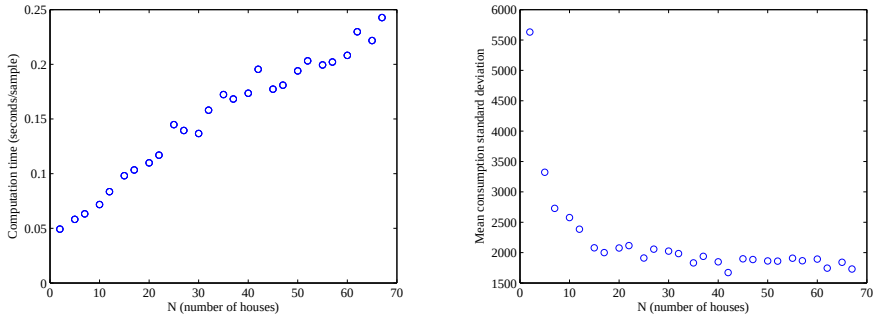


Figure 3.15: *The system seems to be memoryless, meaning that is it independent of initial conditions after a certain time. This is seen as the variations of the indoor air temperatures across the population of houses rapidly decreases. After a couple of days, all indoor temperatures seem to establish a characteristic pattern.*



(a) *Computations run in linear time, which allows for efficient scaling in the number of houses.*

(b) *Decaying standard deviations of the individual house consumption. The mean consumption seems to be representative of the whole population for $N = 20$ houses.*

Figure 3.16: *Computation time and effect of aggregation*

In order to validate the data generated by the simulation framework, a real data set of Danish consumption (obtained from elværksstatistikken) is used for comparison. It comprises a total of have 3234 meterings across 2009. Having access to the part of the consumption coming from households, two time series from January 2009 are compared (Figure 3.17).

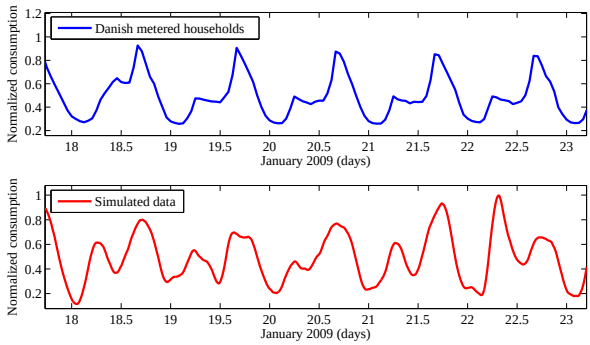


Figure 3.17: *Normalized consumption of 20 simulated houses, smoothed with a Gaussian kernel, against real consumption (also normalized) of Danish households for January 2009. Even though user behaviour was taken from the Olympic Peninsula data set, and that only heating is simulated, the patterns seem to correspond.*

Chapter 4

Identifying consumption structure and price responsiveness

Equipped with a strong data foundation, understanding its structure is essential to model the reaction of consumption to a varying electricity price. That response, being the cornerstone of the whole project, is based on several variables. But which are the most important? To which extent can the consumption be explained without knowledge of the individual households? How big is the price responsive part? How does this response scale or depend on specific parameters?

4.1 Identifying variables

An important element setting the foundation for the modelling of the aggregated consumption is the right selection of variables that will serve as model inputs. The procedure of identifying the most relevant variables is called variable selection (also known in some fields as feature selection). It is used in many pattern recognition and data mining problems, in order to overcome the fact that models become more and more complex when increasing the number of explanatory variables. It is therefore highly important to select the variables that will yield the best modelling performances.

In order to identify the variables best explaining the electricity consumption, we make use of variable selection methods, which can be subdivided into *wrapper* and *filter* methods. Wrapper methods use a flexible model structure to select the best set of input variables minimizing a certain model error. Filter methods on the other hand are independent of an underlying model, but uses a measure of the amount of information in each variable. Wrappers have the drawback of being specifically characterized by a certain type of model and usually have a heavier computational cost. However, their performance is better compared to filter methods that are usually less computationally intensive [26]. Examples of wrapper methods include linear regression models or neural networks; filter methods include the correlation measure, and mutual information.

4.1.1 The forward feature selection procedure

Given a set $\mathbf{X} = (X_1, \ldots, X_D)$ of D potential variables explaining Y, the objective is to order this set by decreasing variable relevance. Two procedures are possible: the forward selection, which starts with an empty set and iteratively adds the most relevant variable until all variable are selected, and the backward selection, starting with the full set and iteratively removing the less relevant variable. Both procedures are independent of the method used (wrapper or filter), and yield an optimal ordering of the variables. As the ordering of the variables requires a significant amount of computation, it becomes advantageous to use the forward selection method as it provides valuable results even though it has not finished ordering the full set.

4.1.2 Using the conditional mean as measure of information

Given a probability space $(\Omega, \mathcal{F}, P)$ with sample space Ω, σ-algebra $\mathcal{F}$ and probability measure P, let

$$L^2 = \left\{X : \mathrm{E}\left\{X^2\right\} < \infty\right\} \tag{4.1}$$

be the set of square integrable random variables (consequently having definite second moments). Furthermore, let $\langle X, Y\rangle$ be the bi-linear form defined by $\langle X, Y\rangle = \mathrm{E}\{XY\}$ for $X, Y \in L^2$. Then $\langle\cdot\rangle$ is an inner product on L^2 with the associated norm $\|X\| = \sqrt{\langle X, X\rangle}$. Under this inner product, the L^2 space is a Hilbert space. One important characteristic of a Hilbert space is that it induces geometry, enabling us to measure the angle between two random variables.

The inner product of two random variables X and Y can be written as

$$\begin{aligned}
\langle X, Y\rangle &= \|X\|\|Y\|\cos(X, Y) \\
\cos(X, Y) &= \frac{\langle X, Y\rangle}{\|X\|\|Y\|} \\
&= \frac{\mathrm{E}\{XY\}}{\sqrt{\mathrm{E}\left\{X^2\right\}\mathrm{E}\left\{Y^2\right\}}} \\
&= \frac{\mathrm{E}\{XY\}}{\sigma_X \sigma_Y}
\end{aligned}$$

One can observe that the cosine of two random variables is the correlation of those variables if they have been centred with zero mean. If the angle between two random variables is small, then the correlation is high, and vice-versa. The correlation coefficient can also be viewed as the cosine of the angle between the two vectors of samples drawn from the two random variables X and Y. This also means that the correlation is minimum when those two vectors are orthogonal, and that the correlation is maximum when the two vectors X and Y are co-linear, meaning that there exists a real k such that $Y = kX$. This implies that the correlation measure is maximized when the relation is *linear*, leaving non-linear relationships troublesome to detect.

An alternative idea is to investigate the amount of information gained by observing X when trying to explain or predict Y. In the task of predicting Y *without* observing X, one should minimize the mean-square prediction error $\mathrm{E}\left\{(Y - \hat{Y})^2\right\}$, simply yielding $\mathrm{E}\{Y\}$ as optimal predictor. However, when an observation of X is available, then the predictor $\hat{Y}$ depends on X. Such a predictor can be written

as $\hat{Y} = g(X)$. It can be shown [12] that a unique optimal predictor minimizing $\mathrm{E}\left\{(Y - g(X))^2\right\}$ exists in the nature of the conditional expectation $\mathrm{E}\{Y|X\}$.

The amount of information gained by observing X can then be defined as the reduction in the mean square prediction error

$$J(X,Y) = \frac{\mathrm{E}\left\{(Y - \mathrm{E}\{Y\})^2\right\} - \mathrm{E}\left\{(Y - \mathrm{E}\{Y|X\})^2\right\}}{\mathrm{E}\left\{(Y - \mathrm{E}\{Y\})^2\right\}} = \frac{\mathrm{Var}\{Y\} - \mathrm{Var}\{Y|X\}}{\mathrm{Var}\{Y\}} \tag{4.2}$$

If X and Y are independent, then $\mathrm{E}\{Y|X\} = \mathrm{E}\{Y\}$ implying $J(X,Y) = 0$ (no new information). On the other hand, if X contains all the information from Y, then $\mathrm{E}\{Y|X\} = Y$ implying $J(X,Y) = 1$. One should note that this measure J is the same as the coefficient of determination R^2, commonly used in statistics.

This motivates the need for methods approximating the conditional expectation function which is a probability density function.

Kernel smoothing

When a probability density function is to be approximated, the simplest method is to use the histogram approximation, where the sample space of the conditioned variable is discretized in k bins. The density is then approximated by the fraction of measurements falling into each bin. As the number of bins increases, the approximation gets better. The drawback of this approach relies on its discreteness, often yielding bigger approximation errors at the boundary of the bins.

An other solution is to use a Parzen-Rosenblatt window method, also called kernel density estimator [14]. Given a set of N observations x_i drawn from a certain distribution X, the kernel density estimator of the conditional density is defined as

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} K(x - x_i) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \tag{4.3}$$

where $K(\cdot)$ is the called the kernel – a symmetric function that sums to one – and h is the smoothing parameter called the bandwidth. A frequently used kernel is the so-called Gaussian kernel, being the standard normal distribution.

$$K(x) = \frac{1}{2\pi} e^{-\frac{1}{2}x^2} \tag{4.4}$$

Intuitively, one would chose the bandwidth h to be as small as possible, however, there is a trade-off between the bias of the estimator and its variance. Several strategies therefore exist to select the smoothing bandwidth h: number of nearest neighbours, cross-validation, minimizing the L^2 risk function... This generalizes to estimators of joint probabilities of dimensions D by using so-called product kernels, where the multidimensional kernel is simply computed as the product of the kernels in each dimension:

$$\hat{f}(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} \prod_{d=1}^{D} K\left(\frac{\mathbf{x} \cdot \mathbf{e}_d - \mathbf{x}_i \cdot \mathbf{e}_d}{h}\right) \tag{4.5}$$

where $\mathbf{x} \cdot \mathbf{e}_d$ is the d-th component of $\mathbf{x}$ (by projection on the basis vector $\mathbf{e}_d$).

For the set of N observations (x_i, y_i) drawn from X and Y, our goal is to compute the conditional expectation

$$\mathrm{E}\{Y|X\} = \int y f(y|x)\, dy = \int y \frac{f(y,x)}{f(x)}\, \mathrm{d}y \tag{4.6}$$

One can use the kernel density estimator to approximate $f(y,x)$ and $f(x)$ in the previous expression, yielding the Nadaraya-Watson estimator $\hat{m}(x)$ of the conditional expectation density $\mathrm{E}\{Y|X\}$

$$\hat{m}(x) = \frac{\sum_{i=1}^{N} y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right)} \tag{4.7}$$

Conditioning on a multivariate random variable $\mathbf{X}$ of dimension D is the same as conditioning on every component of that multivariate variable

$$\begin{aligned} \mathrm{E}\{Y|\mathbf{X}\} = \mathrm{E}\{Y|X_1, X_2, \ldots, X_D\} &= \int y f(y|x_1, x_2, \ldots, x_D)\, dy \\ &= \int y \frac{f(y, x_1, x_2, \ldots, x_D)}{f(x_1, x_2, \ldots, x_D)}\, dy \end{aligned} \tag{4.8}$$

Using the kernel density estimators, the Nadaraya-Watson estimator in D dimensions is

$$\hat{m}(\mathbf{x}) = \hat{m}(x_1, x_2, \ldots, x_D) = \frac{\sum_{i=1}^{N} y_i \prod_{d=1}^{D} K\left(\frac{x_d - x_{1,i}}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{x_d - x_{d,i}}{h}\right)} = \frac{\sum_{i=1}^{N} y_i \mathbf{K}\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}{\sum_{i=1}^{N} \mathbf{K}\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)} \tag{4.9}$$

Note that a generalization to different bandwidths $\mathbf{h} = \{h_1, h_2, \ldots, h_D\}$ for each dimension can easily be made, yielding the final estimator for the conditional

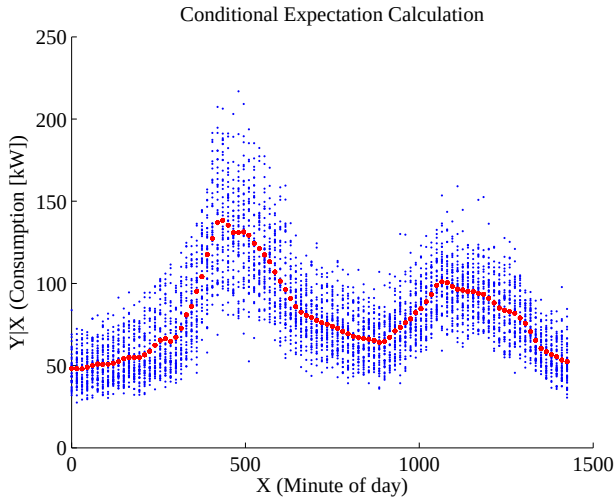


Figure 4.1: *Estimation of the relationship between the variables minute of day and consumption in the Olympic Peninsula dataset by estimating the conditional mean with a variable bandwidth selected upon the $k = \sqrt{N} = 20$th nearest neighbour, yielding the measure $J(X, Y) = 64.3\%$.*

expectation density

$$\hat{m}(\mathbf{x}) = \frac{\sum_{i=1}^{N} y_i \mathbf{K}_{\mathbf{h}}(\mathbf{x} - \mathbf{x}_i)}{\sum_{i=1}^{N} \mathbf{K}_{\mathbf{h}}(\mathbf{x} - \mathbf{x}_i)} \tag{4.10}$$

where

$$K_{\mathbf{h}}(\mathbf{x}) = \prod_{d=1}^{D} K\left(\frac{\mathbf{x} \cdot \mathbf{e}_d}{\mathbf{h} \cdot \mathbf{e}_d}\right) \tag{4.11}$$

The amount of information in a set $\mathbf{X} = \{X_1, X_2, \ldots, X_D\}$ of random variables that can be used to explain a target variable Y can then be assessed by $J(\mathbf{X}, Y)$ in equation 4.2, requiring equation 4.10 to estimate the conditional expectation density.

Because the choice of bandwidth impacts the amount of smoothing used in the estimator, it has to be carefully chosen. As it is highly unlikely that the variable conditioned on has a constant density function, the bandwidth must be set to vary upon the density of measurements to cope with areas with a low number of observations. For example, regions of very high prices occur at very few times, motivating the use of a variable bandwidth. A simple approach is to have

the bandwidth be proportional to the distance to the k-th nearest neighbour. Because the kernel is symmetric, the bandwidth is then defined as half this distance. Generalizing in dimensions D, the bandwidth $\mathbf{h} = \{h_1, h_2, \ldots, h_D\}$ to be used at each point $\mathbf{x}_i$ is defined as

$$\forall d = 1, \ldots, D,\ h_d = |\mathbf{x}_{i,k} \cdot \mathbf{e}_d - \mathbf{x}_i \cdot \mathbf{e}_d| \tag{4.12}$$

where $\mathbf{x}_{i,k}$ is $\mathbf{x}_i$'s k-th nearest neighbour.

4.1.3 Using a neural network to assess variable dependencies

The target variable Y is now represented through a functional relation between the output Y and the D input variables $\mathbf{X} = (X_1, \ldots, X_D)$. This functional relation can be obtained by training of a feed–forward two–layer network for example, with M number of hidden units (Figure 4.2).

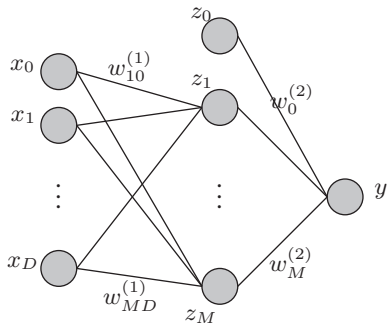


Figure 4.2: *Structure of a feed–forward two–layer neural network with one output, showing the structural relation between an output observation y, and a D-dimensional observation $x_1, \ldots, x_D$. Note that x_0 is the bias of the first layer and z_0 is the bias of the second layer.*

Each input sample $\mathbf{x}$, and a bias term x_0 (of the first layer), are fed through a first layer of weights, $w_{ji}^{(1)}$ to form new inputs a_j fed to the *hidden* units, which transform the inputs using a non–linear *activation function*, $h(a_j)$, where $j = 1, \ldots, M$. Subsequently, the resulting outputs z_j plus a second bias, z_0, are fed through the second layer of weights, $w_j^{(2)}$ to form the output y. Putting this structure for a regression model in mathematical terms and using the hyperbolic

tangent as activation function h yields the non–linear input–output relation

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M} w_j^{(2)} \tanh\left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i\right) \tag{4.13}$$

where the bias terms of the first and second layer were absorbed by setting the variables $x_0 = 1$ and $z_0 = 1$, respectively.

The network parameters are determined using the *backpropagation* learning rule in order to minimize the sum–of–squares error function with respect to the parameters $\mathbf{w}$ over a training set $\{\mathbf{x}_i, y_i\}_{i=1}^N$

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^{N} ||y(\mathbf{x}_i, \mathbf{w}) - y_i||^2 \tag{4.14}$$

Considering identical, independent and Gaussian distributed errors, the optimal parameters $\mathbf{w}$ can be estimated by minimizing the squared error function through a non-linear solver [3]. Note, that equation (4.14) is equivalent to maximizing the maximum–likelihood function of observing the data set $\{\mathbf{x}_i; y_i\}_{i=1}^N$.

The optimum balance between over- and under-fitting of the network to the available data set, i. e. finding its best generalization abilities, can further be improved by putting a quadratic regularization term to (4.14),

$$\tilde{E}(\mathbf{w}) = E(\mathbf{w}) + \frac{1}{2}\lambda||\mathbf{w}||^2. \tag{4.15}$$

The number of hidden units M from Equation (4.13) and the optimal values for λ in (4.15) are determined using K–fold cross validation of the data points ensuring the best possible net. The technique of K–fold cross validation involves taking the available data and partitioning it into K groups, e. g. of equal size. Then $K - 1$ of the groups are used to train a particular network that is subsequently evaluated by means of the remaining group, i. e. its test error. This procedure is then repeated for all K possible choices for the left-out group and the performances from the runs (test errors) are then averaged. The best performing network among all cross-validation procedures is selected and assumed to provide good generalization.

Having found the optimal set of parameters, the estimated output $\hat{y}$ is obtained by simply feeding inputs $\mathbf{x}$ through the trained network.

The network training can now be embedded into the forward selection method, and variable dependency can be assessed by using as measure $J(\mathbf{X}, y)$ the reduc-

tion in variance previously used in Equation (4.2). Its discrete estimator can be written as

$$J(\mathbf{X}, Y) = \frac{\sum_i (y_i - \bar{y})^2 - \sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \bar{y})^2} \tag{4.16}$$

where $\bar{y}$ denotes the mean of the target variable.

4.1.4 Min–Redundancy Max–Relevance (mRMR) with mutual information

Selecting variables of interest means maximizing a certain dependency measure. This is called the max–dependency criterion. However, a combination of variables individually very good at describing the target variable is not necessary better as those variables could be redundant. Therefore, variables should be selected with the criterion of having minimum redundancy among all other selected variables, while still showing high relevance to the target. This is called the Min–Redundancy Max–Relevance (mRMR) criterion which has found to be more powerful than the single max–dependency [22]. This method still relies on a measure of dependency, which could be based on a wrapper method as described previously. However, let us investigate the Mutual Information measure. As it does not base itself on a model, it is a *filter* method (like the correlation measure for example). The Mutual Information between a random variable X and a target variable Y is expressed as

$$I(X, Y) = \int \int p(X, Y) \log \frac{p(X, Y)}{p(X)p(Y)} dxdy \tag{4.17}$$

where the marginal and joint probability density functions are denoted by $p(\cdot)$. The concept of mutual information descends from information theory and is based on the *entropy* of a random variable. Shannon expressed the entropy of a random variable as the expected value of its information content [3]. The information content only depends on the probability of the random variable: the smaller its probability, the larger the information content associated with receiving the information that the event indeed occurred. By the fact that, by definition, the measure of information is positive and additive, the information content of a random variable X is expressed as the negative logarithm of the probability distribution function $p(X)$. The entropy $H(X)$ of that random variable is then the expected value of the information content, such that

$$H(X) = -\int \log(p(X)) \tag{4.18}$$

Mutual Information measures the entropy shared by the two variables (also called *joint* entropy), such that $I(X,Y) = H(X,Y) - H(X|Y) - H(Y|X)$, yielding Equation (4.17). This measure can then be used as an alternative to the reduction of variance measure $J(\mathbf{X}, Y)$ from Equation (4.2). However, an improved criterion from the max–dependency (that would just maximize this measure) is the mRMR. It has been shown in [22] that if the the max–dependency criterion is used together with the min–redundancy (mRMR), then using pairwise mutual information to select the most important variables is equivalent to using the multivariate equivalent. The mRMR score for a variable X_i that is not picked yet is then computed as

$$\text{mRMR}_{X_i \in R} = I(X_i, Y) - \frac{1}{s} \sum_{j \in S} I(X_i, X_j), \tag{4.19}$$

where S and R denote sets of s already selected and r remaining variables, respectively. Finally, that variable among the r scores is chosen to be included yielding the maximum mRMR score.

A crucial part in computing the mRMR scores is a proper bivariate probability density estimation in Equation. (4.17). This is carried out using a two–dimensional kernel density estimator with automatic bandwidth selection[1].

[1] Implementation taken from Zdravko Botev (2009). Available under `http://www.maths.uq.edu.au/~botev`

4.1.5 Results of the variable selection

The three variable selection methods are now applied to the two datasets produced in Sections 2.2 and 3.2 in order to assess the importance of each variable. Another dimension is added as each variable is delayed (lagged), up to one day, enabling us to identify within-a-day dynamics. The selection methods are run on two sets of variables: The external set of observables, and the set of external *and* internal variables:

External variables	Internal variables
Price	Standardized price
Minute of day (not lagged)	Inside air temperature
Weekday (not lagged)	Heating setpoint
Weekend boolean (not lagged)	Heating setpoint adjustment
Outside temperature	
Sun global irradiance	
Sun diffuse irradiance (Olympic Peninsula only)	
Sun direct irradiance (Olympic Peninsula only)	
Outside humidity (Olympic Peninsula only)	
Outside dewpoint (Olympic Peninsula only)	
Outside wind speed (Olympic Peninsula only)	
Outside wind direction (Olympic Peninsula only)	
Barometer (Olympic Peninsula only)	

A summary of the results is presented here. Detailed scores are listed in Appendix B.

Olympic Peninsula

The strongest dependency is on the time of the day, as all methods pick either the MinuteOfDay or the heating setpoint as their first variable (note that the heating setpoint is a representation of the time of the day since it is triggered by occupancy modes). The second most important variable is the outside dewpoint, which surpasses the outside temperature by far. The lagged heating setpoint are selected throughout the procedures, indicating a importance of its dynamics. However, the dependency on weekday and weekend variables is less systematic, as only some of the methods select them, in very different orders. The sun irrandiance is even more problematic, as it is only selected by the neural network, and ranked #5. Price responsivity is only identified by the kernel as the 12th

variable. When the internal standardized price is included, is is ranked #4 by the kernel and #8 by the neural network (the mRMR does not detect any price responsivity).

Simulated data

From the outside, the simulated system is very price responsive: most of the variables selected are lagged versions of the price. Furthermore, the kernel selects 8 lagged prices and the mRMR selects 14 (out of 15). All three methods show the strongest variable to be the non-lagged price, closely followed by the time of the day, the outside temperature and the sun irradiance. Performances are quite limited when using only external variables. However, adding internal variables drastically increases performances. Scores go as high as 80% after a couple of variables, and maximum scores are >90% with 10 variables. Throughout the three methods, the first variable selected is the indoor temperature, explaining 45% of variance. The kernel and the mRMR select as second variable the indoor temperature again, but delayed with 5 minutes (one sample). This clearly represents the first order differential equation driving the building dynamics. The standardized price is then quickly selected, with numerous lagged versions of the indoor temperature, stating the strong dependence on its dynamics.

4.2 Heating consumption structure

The variable selection used in the previous section revealed a strong dependency of the consumption on the heating setpoint. We will therefore focus on understanding the structure of the part of the consumption that is related to heating. By building the simulation framework close to the Olympic Peninsula experiment, we identified relationships between variables, structured in Figure 4.3. This structure is valid for both datasets, and all subsequent analysis will therefore be treated on both datasets simultaneously.

For each house, the heating consumption is determined by the deviation of the inside air temperature T_a to a certain reference (being the heating setpoint). If the inside air temperature is under the reference such that heating is required, then an according consumption is observed. This triggers a feedback of heat Q from the heating system to the building, which will in return affect the inside air temperature, the latter also being affected by outside weather conditions.

Given a heating setpoint, the heating consumption is then a result of an interplay between the quantity of heating applied Q and the inside air temperature T_a, the latter also being affected by weather conditions (upper part of Figure 4.3).

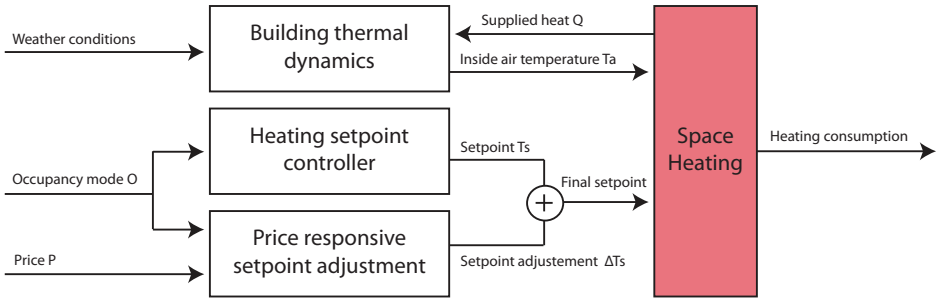


Figure 4.3: *Block diagram of the heating consumption. External variables are entering from the left, transformed into internal (hidden) variables, before being transferred into the output consumption.*

The temperature setpoint on the other hand can be decomposed in two parts: a price responsive adjustment dT_s and a non-price sensitive part T_s. The non-price sensitive part of the setpoint is determined by the occupancy modes: Each mode triggered a specific setpoint. On the other hand, the price responsive adjustment originates from the standardized price $\hat{p}$, the price sensitivity k and the associated bounds of comfort (maximum and minimum acceptable temperature setpoints). The price responsive adjustment therefore depends on the occupancy modes (through the comfort settings), and on the received price (through the standardized price).

The overall structure is generalizable to other systems than heating. As an example, the consumption of price-responsive fridges and freezers, as implemented in Chapter 3, follows the same structure as it is based on the same principles of lowering/increasing the setpoint depending on a price signal.

4.2.1 Aggregation

The model structure for an *individual* consumer's heating system has been described. However, how can a population of different households be described? In other terms, what does the *aggregated* system look like?

The heating consumption of one house does not affect the consumption of others.

Even though different households depend on the same external variables (they all react to the same prices and to some extent to the same weather conditions), their dynamics are independent of each other. It seems therefore reasonable to investigate to which extent the mean behaviour of households is representative of the population, given that we consider enough households, e. g. that the law of big numbers prevail.

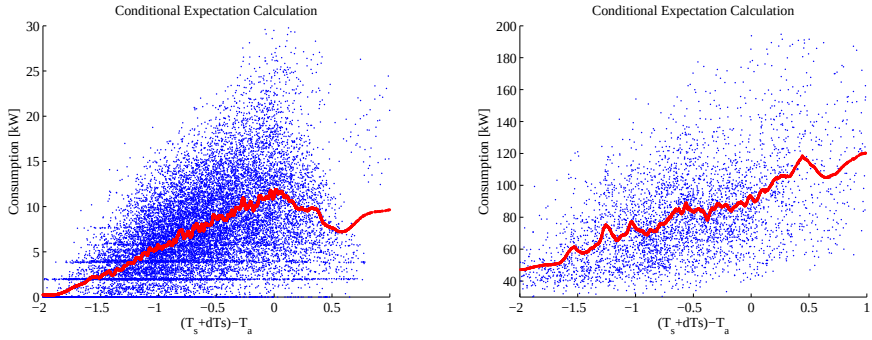
For the Olympic Peninsula data there is no other choice as to assume this fact since only aggregated time series of consumption, comfort and external conditions were given. But since the mean of each variable's time evolution follows a clear pattern and the deviations from it are in a passable range, it is reasonable to assume a similar behaviour for the mean. For the simulated data this assumption holds likewise, as investigated in Section 3.2 when preparing a simulated data set. It is also important to mention that the heating consumption in the Olympic Peninsula was blurred by other devices' consumption since they could not be separated from the aggregated consumption data.

We will therefore approximate the aggregated system by the „mean system" of Figure 4.3 using the mean of all variables used as input and output (except the consumption which will be the sum of all individual consumptions). The strategy is to investigate and understand each subsystem in order to build an external description of the overall system, as we don't have access to the internal (hidden) variables.

4.2.2 Space heating

The space heating system can generally be thought of a simple controller that turns on or off the heating device according to whether the heat supply is needed or not. One should note that this internal process of supplying heat to the building results in an internal feedback, which makes the overall system non-linear.

Inputs to this subsystem are the indoor temperature T_a, the temperature setpoint $T_s + dT_s$ (consisting of both price-responsive and non price-responsive parts). The consumption c is the output. In a single house, heating is turned on if $T_a < T_s + dT_s - T_{hys}$. Figure 4.4 indicates a clear trend between the difference of the inputs to the system and the resulting consumption for both the data sets. This relation is static, and it is not expected that any dynamics are introduced here as the decision to turn on heating depends on the present values of T_a and $T_s + dT_s$.



(a) *Simulated data.* **(b)** *Olympic Peninsula.*

Figure 4.4: *Consumption as a function of the difference between heating needs and current indoor air temperature,* $(T_s + dT_s) - T_a$.

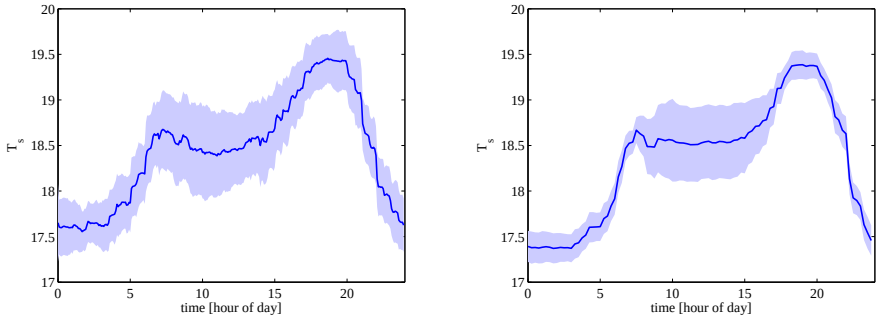
4.2.3 Building thermal model

The indoor air temperature T_a used as input to the space heating is determined by a system having as input external weather conditions (outside temperature T_o, sun irrandiance S) and the amount of heat supplied Q as a feedback from the space heating system. As explained in Section 3.1.2, if no heating is applied, the transfer function of the thermal model for a building can be seen as a lowpass filter with a corresponding phase shift, i. e. the fast fluctuations of the inputs are attenuated and time delayed (see Figure 3.4).

4.2.4 Heating setpoint controller

The setpoint T_s is determined by a system having the occupancy modes as input, triggering a preassigned temperature setpoint as output, reflecting the time varying human comfort needs of the household.

The occupancy modes are found to be triggered according to the time of the day and to have slightly different structures for weekends or weekdays (see Figure 2.5). It could be assumed that they also depend on holidays or other punctual events. Therefore, the heating setpoint controller system could be assumed to have as only input the time of the day, and an indicator function separating weekdays and weekends.



(a) *Simulated data.* **(b)** *Olympic Peninsula.*

Figure 4.5: *Heating set point as a function of time of the day.*

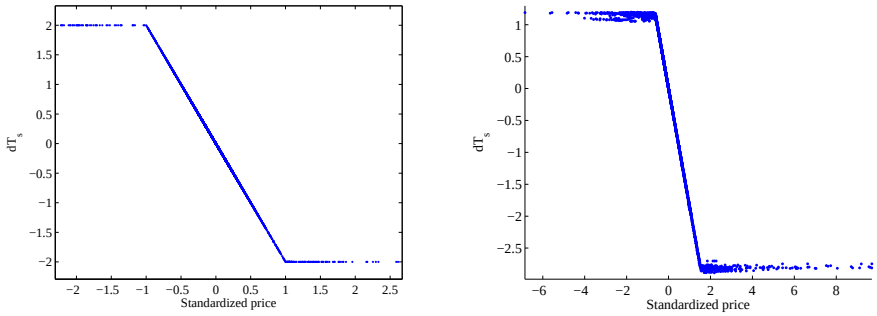
The average setpoint is found to be representative of the population (Figure 4.5). The time of the day seems to be a very good explanatory variable to describe this mean setpoint.

Note that a simple heating system controller is assumed here, based on a simple on/off switch.

4.2.5 Price responsive setpoint adjustment

In order to fully describe the final setpoint given as input to the space heating, we look at the price responsive setpoint adjustment dT_s, affected by the received price and the triggered occupancy mode. The price is firstly standardized as described in Section 1.2 and subsequently transformed into a set point offset according to the price sensitiveness and comfort bounds, which are also triggered by the occupancy modes. In both data sets, the price sensitivity and comfort bounds were found to stay close to constant. Consequently, the price responsive setpoint adjustments appears almost constant for a given price, as shown in Figure 4.6.

As in the heating setpoint controller, the effect of the occupancy mode can be quite well approximated by a dependency on the time of the day.



(a) *Simulated data.* **(b)** *Olympic Peninsula.*

Figure 4.6: *Setpoint adjustment as a function of standardized price.*

4.2.6 Conclusions

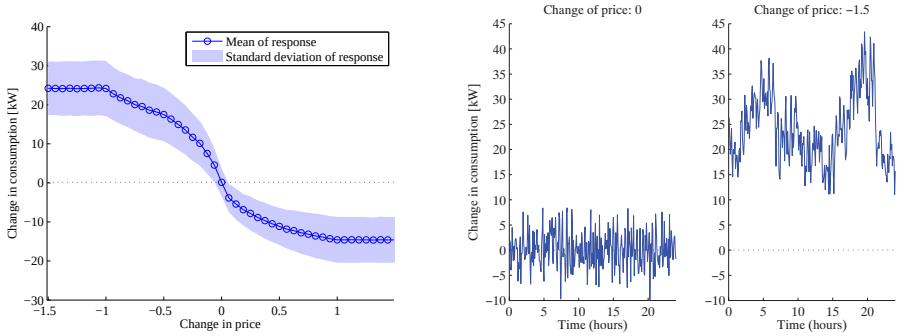
The interaction between the heating system and the building can be considered as a hidden process since T_a and Q will most likely not be observable at an aggregated level in a real-life implementation. The heating setpoint T_s, the standardized price ρ together with the associated price reference $\bar{p}$ are likewise hidden. The challenge is therefore to conceive a model capable of describing accurately enough the aggregated consumption, making use of only external variables.

4.3 Price responsivity of heating systems

Having as basis the structure of the consumption, we move on to the task of extracting the price response and investigating its structure and dependencies. For this purpose we firstly draw benefit of the simulation framework by *measuring* the change in consumption after a change of price, by assuming that all other variables vary so slowly that they can be assumed constant. Subsequently, we take a closer look at statistical modelling in order to identify a mathematical description in terms of an impulse response of consumption to price. Finally, price responsiveness is considered relating to dependencies on other variables. Note that although we focus on using only external variables, internal variables are included here when necessary, in order to properly identify and extract the price responsive structure.

4.3.1 Simulating the price response

The real response of the system can be investigated by exciting the (simulated) system with different prices and thereafter measuring its response. This is effectively done by using as initial system two days with 20 houses. Two days are used in order to ensure convergence from the initial conditions. The final state of the system is then extracted, the price is modified, and one sample is simulated with this modified state. As a reference case, no price change is performed. The resulting consumption is measured, assuming that all internal variables have not had the time to change significantly during one sample. Because the price during the two first days was kept constant to 1, the change in price is actually the same as the standardized price ρ.



(a) *Consumption change after a change in price. Standard deviation computed across one day.*

(b) *Consumption change across the day for a fixed change of price. The responsivity of the system depends on the time of the day, and is maximized during peak hours.*

Figure 4.7: *Simulated one-step consumption response for various changes in price. Note that the system was not exposed to any prices before.*

A non-linear one-step price response is observed, with responsivity saturations for price changes higher than 1 and lower than -1. The response also depends on the time of the day as seen in Figure 4.7b. Maximum responsivity is achieved during periods of high heating demands, corresponding to the morning and evenings peaks.

This is only the one-step price response, independent of previous price changes (as none was applied before). A more systematic approach needs to investigate how the response evolves over time (how long does a price change affect the system?) and how this price response depends on previous price changes (how

much flexibility is left depending on previous price changes).

4.3.2 Modelling the price response with a linear model

The objective in now to identify a statistical model sufficiently precise to describe the response of consumption to a change in price. The variable selection methods in Section 4.1 outlined four important variables: price, heating setpoint (or time of the day), outside temperature and sun irradiance. Even though we aim at modelling without knowledge of the internal variables, the heating setpoint is preferred as a first step because introducing the minute of the day would immediately cause a non-linearity (see Figure 4.1).

Modelling is carried out by simultaneously considering the simulated and real data set where especially the outcome of the Olympic Peninsula project (cf Section 2.3) is promising regarding price responsiveness.

It makes sense to start with linear models, simple by nature and well studied. Given certain assumptions, this class of models provide closed-form solutions that can therefore be computed quickly and precisely. The *Auto–Regressive* model with *eXogeneous* inputs (ARX) enables a description of the target variable (the consumption) to linearly depend on previous values of the target and on current and previous values of external variables (price, weather...).

For a single external variable x, the ARX model describing the output y_t can be written as [15]

$$y_t + \phi_1 y_{t-1} + \ldots + \phi_{n_a} y_{t-n_a} = \theta_0 + \theta_1 x_t + \ldots + \theta_{n_b+1} x_{t-n_b} + \epsilon_t, \qquad (4.20)$$

where ϵ_t is a zero mean Gaussian error. By denoting the finite range of past outputs by $\mathbf{y}_{t-1} = (y_{t-1}, \ldots, y_{t-n_a})^T$ and the finite range of past inputs by $\mathbf{x}_t = (x_t, \ldots, x_{t-n_b})^T$, Equation (4.20) can be reformulated as

$$y_t = \begin{pmatrix} -\mathbf{y}_{t-1}^T & 1 & \mathbf{x}_t^T \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\theta} \end{pmatrix} + \epsilon_t. \qquad (4.21)$$

Note that multivariate inputs can be introduced by concatenating the additional dimensions horizontally into Equation (4.21). The range of past output measurements integrated into the model is controlled by the parameter n_a. n_b equivalently controls the amount of past external inputs used.

Vertically stacking all observation vectors into matrices, the ARX model can be

formulated as the linear model

$$\mathbf{y} = \mathbf{X}^T \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\theta} \end{pmatrix} + \boldsymbol{\epsilon} = \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{4.22}$$

The unknown parameters $\boldsymbol{\beta}^T = [\boldsymbol{\phi}^T \ \boldsymbol{\theta}^T]$ are found by using the least squares estimator [15]

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{4.23}$$

Note, that extending the ARX structure in 4.21 to bivariate outputs $y_t^{(1)}$ and $y_t^{(2)}$ can be done by expanding the matrix equations such that

$$\begin{pmatrix} y_t^{(1)} & y_t^{(2)} \end{pmatrix} = \begin{pmatrix} -\mathbf{y}_{t-1}^{(1)T} & -\mathbf{y}_{t-1}^{(2)T} \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi}_{y^{(1)}} & \boldsymbol{\phi}_{y^{(2)}} \\ \boldsymbol{\varphi}_{y^{(1)}} & \boldsymbol{\varphi}_{y^{(2)}} \end{pmatrix} + \mathbf{x}_t^T \begin{pmatrix} \boldsymbol{\theta}_{y^{(1)}}^T & \boldsymbol{\theta}_{y^{(2)}}^T \end{pmatrix} + \boldsymbol{\epsilon}_t \tag{4.24}$$

A special case of the ARX model is the so-called Finite-Impulse-Response (FIR) model, which only incorporates external inputs neglecting any dynamics in the output ($n_a = 0$),

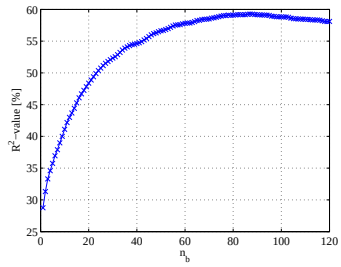
$$y_t = \begin{pmatrix} 1 & \mathbf{x}_t^T \end{pmatrix} \boldsymbol{\theta} + \epsilon_t \tag{4.25}$$

The parameter vector $\boldsymbol{\theta}$, without the first coefficient representing the intercept, can therefore be seen as the impulse response of the system.

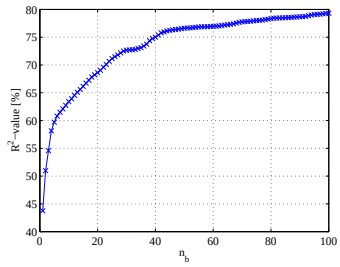
ARX and FIR models are widely studied models in statistics and are very often used as a first start in system identification since they are simple to estimate and evaluate. A basic assumption of these types of models is an underlying stationary process, i. e. the first and second order moments of the modelled processes, mean and variance, do not change over time. Various techniques like detrending or filtering exist in order to make non-stationary time series stationary, see [15]. However, for the system identification part, we used almost stationary periods of data (only one or two months) in order to reduce those problems. Furthermore, one should note that by exciting the system with different prices, the price reference $\bar{p}_t$ (upon which the standardized price ρ_t is computed) changes over time. Even though the response to the standardized price ρ_t is stationary, the response to the external price p_t isn't, as the price reference $\bar{p}_t$ changes. We have therefore taken care to generate and model a data set which have prices yielding a fairly constant price reference.

Determining the best number of lags for each input variable was done by evaluating each model performance for different lag values (Figure 4.8). On the simulated data set, enough data points were available such that it was possible to assess the model performance on a test set, after having been estimated on a

training set. Note that one could exploit more advanced model assessment and selection techniques as described in [15].



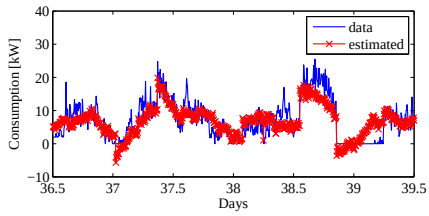
(a) *Simulated data.*



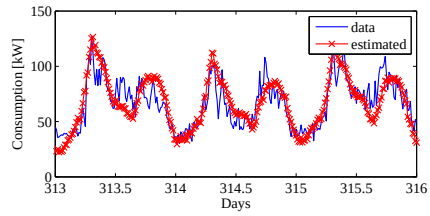
(b) *Olympic Peninsula.*

Figure 4.8: *Determination of best number of lags for FIR model in* (4.25) *by means of percentage of reduction in variance. Note that because the performance on the simulated data was evaluated on a test set, effects of overfitting appear for* $n_b > 80$.

First, we examined the performance of an FIR model on both data sets by assessing the reduction of variance between this model and a constant model (the mean of the output). $n_b = 80$ lags (6.7 hours) were selected for the simulated data and $n_b = 40$ lags (10 hours) for the Olympic Peninsula, selected as the value for which an increase in lags did not significantly increase the coefficient of determination R^2. A period of the time series for both estimations are shown in Figure 4.9. It can be observed in both plots that the mean behaviour is better captured than the high fluctuations of the consumption. Changing to an ARX model, i. e. dynamics in the consumption are added, did not yield any significant improvement.



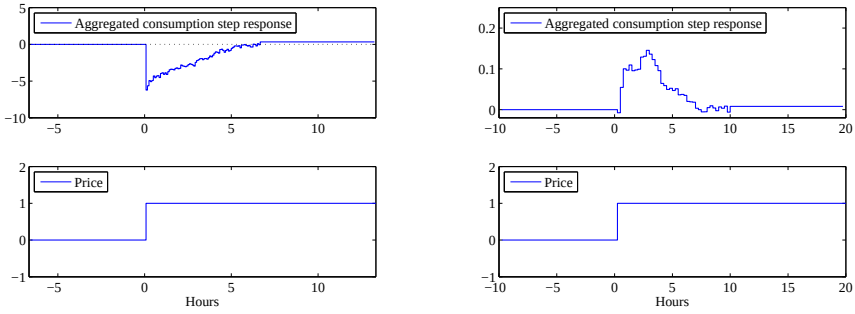
(a) *Simulated data*



(b) *Olympic Peninsula*

Figure 4.9: *Exemplary true and estimated time series using the FIR model in* (4.25) *with 80 lags selected for the simulated and 40 lags for the real data set.*

The price responsive parts of the models can now be extracted. The response of the consumption to a step of price is therefore investigated by convolving the models' impulse response of consumption to price with a step function of one price unit. The resulting step responses for both data sets are depicted in Figure 4.10.

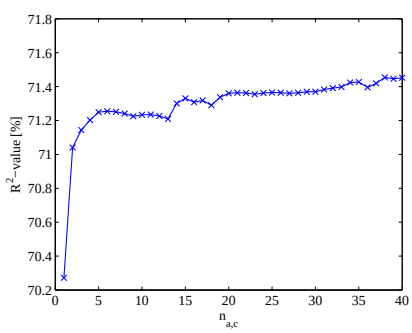


(a) *Simulated data* **(b)** *Olympic Peninsula*

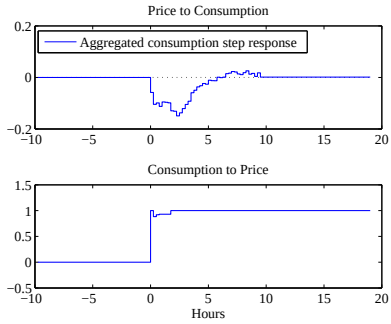
Figure 4.10: *Aggregated consumption step response. A change of price influences the simulated consumption during approximately 5 hours. The counter-intuitive response of the Olympic Peninsula shows a necessity of taking into account the closed loop relation between price and consumption.*

Considering the simulated data, an increment in price immediately reduces the consumption. In the Olympic Peninsula however, the opposite is observed. This comes from the fact that price and consumption are affecting each others, whereas in the simulation framework, an increase in consumption did not affect the broadcasted price. This mutual influence of price and consumption states describes a *closed loop* system.

By making use of an ARX having as additional output the price (meaning price is not anymore an external input), closed loop models can be described with an ARX structure according to Equation (4.24). As we already found $n_b = 40$ sufficient for the external variables in the FIR, the number of lags $n_{a,p}$ and $n_{a,c}$ necessary to describe the price and consumption dynamics, respectively, have to be chosen. The number of lags reflecting the price dynamics, $n_{a,p}$, is chosen to be 40 (10 hours) enabling a proper representation of the impulse response of consumption to price. On the other hand, $n_{a,c}$ is selected using the same method as previously described for the FIR model. $n_{a,c} = 4$ seems to be sufficient (Figure 4.11a) since higher values do not significantly improve the model performance.



(a) *Determintation of best number of lags for the consumption in the multi-output ARX model in* (4.24) *by means of percentage of reduction in variance.*



(b) *Step response of the decoupled system. An increase in price causes a decrease in consumption that lasts approximately 7 hours, and an increase of consumption causes an increase of price. The low scaling for the response in the Olympic Peninsula experiment is due to an overall high level of prices.*

Figure 4.11

The responses of consumption after a step of price (the feedforward) and of the price after a step of consumption (the feedback) can then be extracted using the same method as before (Figure 4.11b). A reasonable response of the consumption to an increase in price can now be observed. Furthermore, it is also observed that an increase in consumption has as consequence a rise in price.

Modelling with internal variables turned out to be more problematic than helpful. Including the indoor temperature blurred out the effect of price, as the indoor temperature already depends on the price. This problems could have been solved by adding this internal variable as output in an ARX, but at a high cost of having the number of parameters increase accordingly.

In that sense, the variables used here are judged adequate in order to explain the price responsivity of the system.

4.3.3 Dependencies on other variables

When modelling the price step response for the data of the Olympic Peninsula experiment, we found the response to vary with the different data periods that

have been chosen for statistical modelling (see Table 2.1).

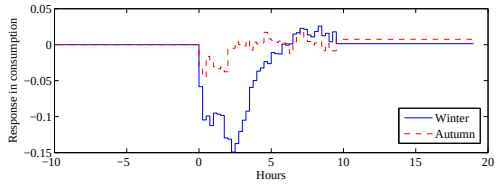
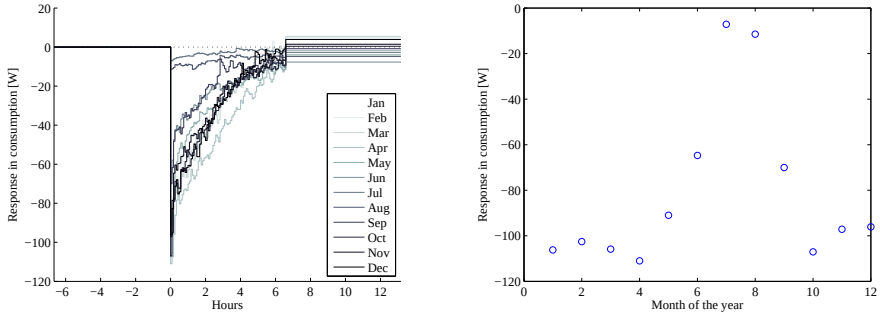


Figure 4.12: *Price step response when different seasons of the Olympic Peninsula project are considered.*

This observation motivates the investigation of how the step response changes depending on various variables. For this purpose, the simulation framework enables us to change several variables and investigate the changes in step response. Several data sets of length of two months are generated and compared.

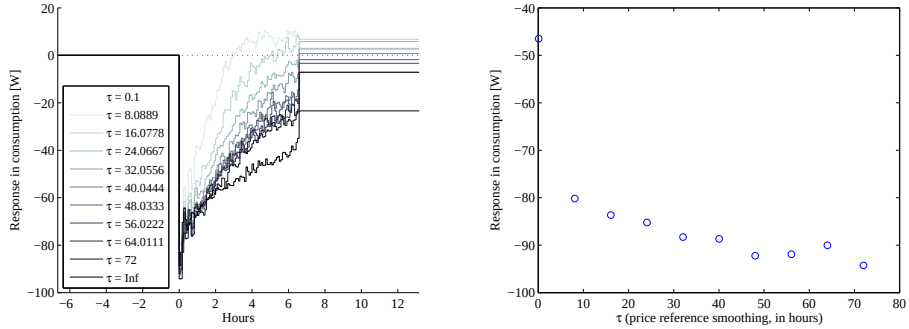


(a) *Step response* **(b)** *Instant response*

Figure 4.13: *Price response when different months are considered. The heating needs are different throughout the year. May-September 2009 are found to be less price-responsive than winter months.*

The time constant τ controls the range of past prices taken into account when computing the reference price $\bar{p}$. For $\tau = 0$, no past price is used to compute the reference price, which is then exactly equal to the received price signal. For τ being infinity, infinitely many old measurement are (theoretically) used to compute the reference price, making it independent of the current price signal.

The smaller τ gets, the more similar the price reference and the received price, lowering the price response. A saturation is observed for $\tau > 30$ hours, where



(a) *Step response* **(b)** *Instant response*

Figure 4.14: *Price response when different time constants of the reference price are considered.*

enough past measurements are taken into account in order to have a stable reference (Figure 4.14). When a constant reference is used (infinite τ), the smoothing effect of the standardized price ρ is completely removed, resulting in a longer response without changing the instantaneous response (Figure 4.15). In that sense, the price response represents both the smoothing effect and the response of the physical system.

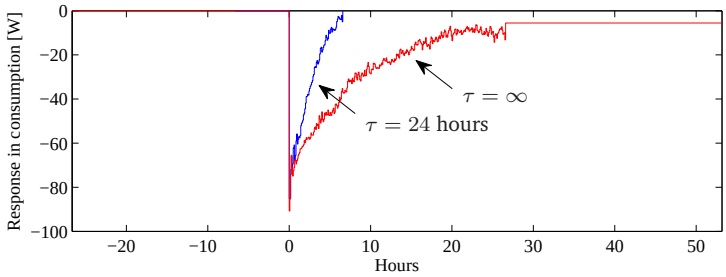
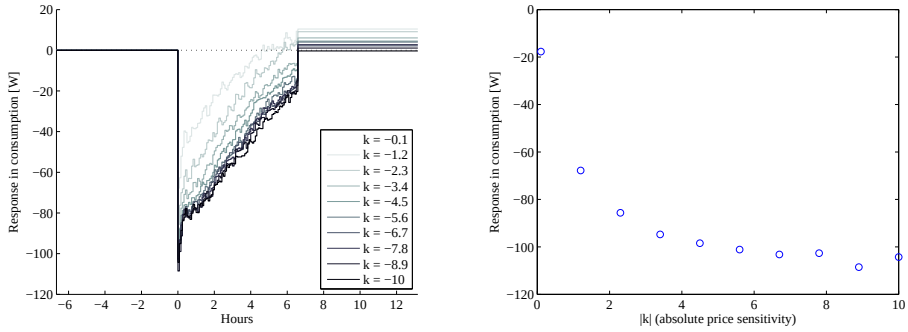


Figure 4.15: *By setting $\tau = \infty$ when computing the price reference, more flexibility is obtained during time, at the cost of increasing the model order ($n_b = 320$). However, the appliances are unable to adapt to slow variations of the price. The time constant should be chosen as big as possible, such that slow variations of the price are taken into account without seriously reducing flexibility.*

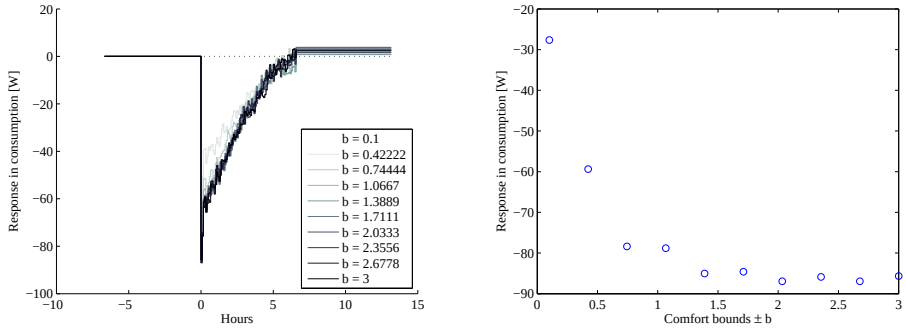
By setting $\tau = \infty$, the effect of price stays approximately 20 hours in the system. The full flexibility potential is therefore not used if the smoothing effect of the

standardized price ρ is present. As the original purpose of this smoothing is to remove low price variations, coming e. g. from monthly or yearly variations, the time constant τ should be chosen as long as possible to maximize the impact of a price change.



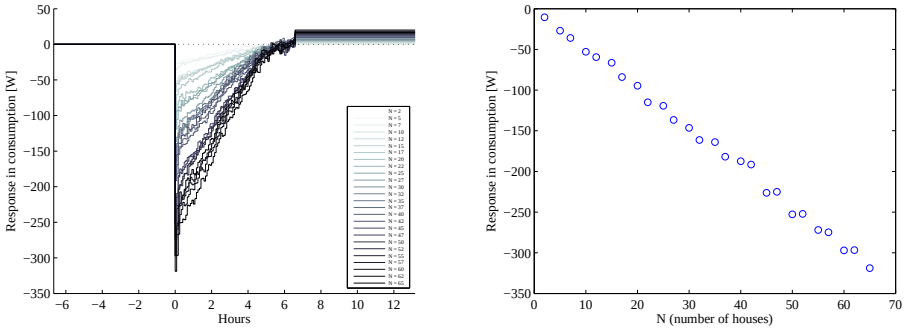
(a) *Step response* **(b)** *Instant response*

Figure 4.16: *Price response when different price sensitivities are considered. All coefficients of the step response increase as price sensitivity is increased. A saturation is observed for sensitivities having an absolute value bigger than 6°C per doubling of price*



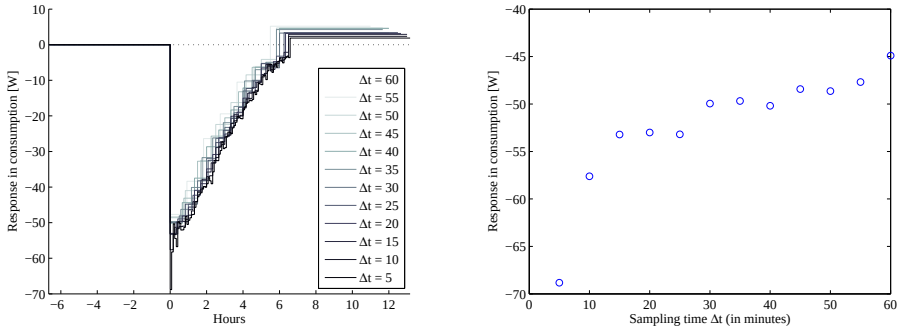
(a) *Step response* **(b)** *Instant response*

Figure 4.17: *Price response when different comfort boundaries are considered. It seems that the comfort bounds affect mostly the immediate response.*



(a) *Step response* **(b)** *Instant response*

Figure 4.18: *The price response linearly scaled with the number of households.*



(a) *Step response* **(b)** *Instant response*

Figure 4.19: *Price response when different sampling times are considered. Sampling times over 15 minutes miss approximatively 25% of the immediate response.*

CHAPTER 5

Forecasting and controlling the flexible consumption

The Olympic Peninsula project controlled the consumption by collecting bids of appliances and matching those against production bids. This yielded a clearance price, broadcasted back to appliances. Because this approach requires appliances to broadcast bids to the market, it therefore needs a robust and secure communication infrastructure.

Our approach seeks at constructing a price generator with the objective of following a certain consumption reference. This controller is actively using the patterns emerging from the aggregated consumption in order to take better decisions. It is completely based on external variables e. g. it assumes no knowledge of the internal variables of each household. But to which degree can we control the prices being generated? How can the generator cope with time-varying consumption patterns? How can we assess the performance of such a generator?

5.1 Concept

The concept enabling the activation of flexible consumption by using a price signal combines three elements: the real system representing an aggregate of price responsive households, a system identification part estimating the price responsivity of the system, and a controller generating the prices needed in order to achieve a given consumption (Figure 5.1).

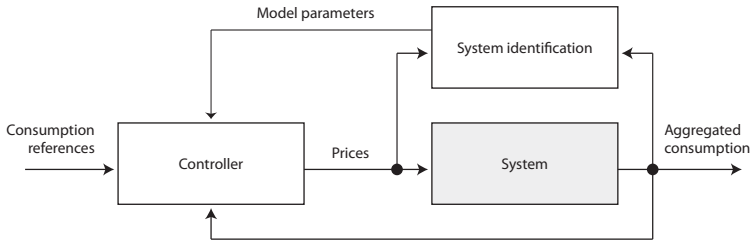


Figure 5.1: *The controller emitting a price signal is able to influence an aggregate of price-responsive households (the system). On-line identification enables the controller to adapt to changes in the system (self-tuning).*

As the number of installed price responsive appliances might vary over time, and as the population itself might change behaviour over time, the system is *time-varying*. Dealing with a system exhibiting time-varying dynamics imposes the use of *adaptive* control. An adaptive controller is formed by combining an on-line system identification, which re-estimates the unknown model parameters at each instant, with a control law based on those on-line estimates. A controller based on a regularly updated model is often referred to as a *self-tuning* controller, since the controller is capable of adjusting itself to changes in the system. Of course, this ability is highly dependent on the robustness and convergence properties of the processes involved, see [13].

External influences such as e. g. outside temperature have also to be taken into account in the system identification, as the real system is affected by them. Current and past values can be measured, but in order to derive a proper control law, forecasts must be provided. Additional uncertainties are introduced in the process of forecasting. Providing reliable forecasts is a discipline in itself. In that light, we have assumed perfect knowledge of the external influences. One the other hand, as internal variables are not accessible from the system identification block, the model from Section 4.3.2 has to be adapted such that it *estimates* internal variables, such as the heating temperature setpoint.

5.2 Forecasting the consumption

5.2.1 Forecasting theory

In this Section we access the performance of the models obtained for both the simulated and Olympic Peninsula data set in Section 4.3.2 in terms of their forecasting capabilities. Especially when used later in a controller, the prediction ability of a model is a crucial factor regarding the overall control performance. The task of forecasting the consumption is also sometimes referred to as *prediction*.

Let us denote $\mathbf{X_t}$ as the matrix containing a finite range of lagged prices $\mathbf{p}_t$ and lagged external variables $\hat{\mathbf{Z}}_t$ up to time t, such that $\mathbf{X_t} = \begin{pmatrix}\hat{\mathbf{Z}}_t & \mathbf{p}_t\end{pmatrix}$. A model f describing the consumption by means of own consumption dynamics and external lagged inputs can be written in a very general form

$$\begin{aligned} c_t &= f\left(\mathbf{c}_{t-1}, \mathbf{X}_t\right) + \epsilon_t && \text{(5.1a)} \\ &= f\left(c_{t-1}, \ldots, c_{t-n_a}, \mathbf{x}_t, \ldots, \mathbf{x}_{t-n_b+1}\right) + \epsilon_t && \text{(5.1b)} \end{aligned}$$

Note that for $n_a = 0$, the model is independent of past consumption values.

By assuming that we want to predict by minimizing the squared error between predictor and true value, the optimal one-step prediction is expressed as the expectation of the consumption conditioned on past measurements up to time, t [15].

$$\begin{aligned} \hat{c}_{t+1|t} &\equiv \mathbb{E}\left\{c_{t+1} \mid \mathbf{c}_t, \mathbf{X}_t\right\} && \text{(5.2a)} \\ &= \mathbb{E}\left\{f\left(\mathbf{c}_t, \mathbf{X}_{t+1}\right) + \epsilon_{t+1} \mid \mathbf{c}_t, \mathbf{X}_t\right\} && \text{(5.2b)} \\ &= f\left(\mathbf{c}_t, \mathbf{X}_{t+1}\right) + \mathbb{E}\left\{\epsilon_{t+1|t}\right\}, && \text{(5.2c)} \end{aligned}$$

where $\mathbb{E}\left\{\epsilon_{t+1|t}\right\}$ can be seen as a bias correction term of the predictor. Note that this term vanishes under the assumption that the model error ϵ_t is i.i.d. Gaussian with zero mean.

The conditional variance, on the other hand, describes the uncertainty associated with the prediction:

$$\begin{aligned} \mathrm{Var}\left\{c_{t+1} \mid \mathbf{c}_t, \mathbf{X}_t\right\} &= \mathrm{Var}\left\{f\left(\mathbf{c}_t, \mathbf{X}_{t+1}\right) + \epsilon_{t+1} \mid \mathbf{c}_t, \mathbf{X}_t\right\} && \text{(5.3a)} \\ &= \mathrm{Var}\left\{\epsilon_{t+1} \mid \mathbf{c}_t, \mathbf{X}_t\right\} && \text{(5.3b)} \\ &= \mathrm{Var}\left\{\epsilon_{t+1|t}\right\} && \text{(5.3c)} \end{aligned}$$

The k-step ahead predictions and their associated uncertainties are also found to be the conditional mean and variance [15] (conditioned on the time t at which a prediction is made):

$$\begin{aligned} \hat{c}_{t+k|t} &\equiv \mathbb{E}\left\{c_{t+k} \mid \mathbf{c}_t, \mathbf{X}_t\right\} && \text{(5.4a)} \\ &= \mathbb{E}\left\{f\left(\mathbf{c}_{t+k-1}, \mathbf{X}_{t+k}\right) + \epsilon_{t+k} \mid \mathbf{c}_t, \mathbf{X}_t\right\} && \text{(5.4b)} \end{aligned}$$

$$\operatorname{Var}\left\{c_{t+k} \mid \mathbf{c}_t, \mathbf{X}_t\right\} = \operatorname{Var}\left\{f\left(\mathbf{c}_{t+k-1}, \mathbf{X}_{t+k}\right) + \epsilon_{t+k} \mid \mathbf{c}_t, \mathbf{X}_t\right\}. \quad \text{(5.4c)}$$

The model f inside the conditional mean and variance now requires unknown values: consumptions at time indices $t+1$ up to $t+k-1$. Those values can be recursively approximated using the one-step predictor previously established, but most often at the cost of having a growing uncertainty due to the recursive approximation error.

However, we can prevent the accumulation of errors by selecting a model having $n_a = 0$, i. e. that the one-step predictor is independent of past consumption values. This results in a predictor which only depends on external variables up to the prediction horizon.

5.2.2 Forecasting the consumption using an FIR model

For the linear FIR model found in Section 4.3.2, described by the parameter vector $\boldsymbol{\theta}$, the k-step prediction from Equation (5.2c) simplifies to [15]

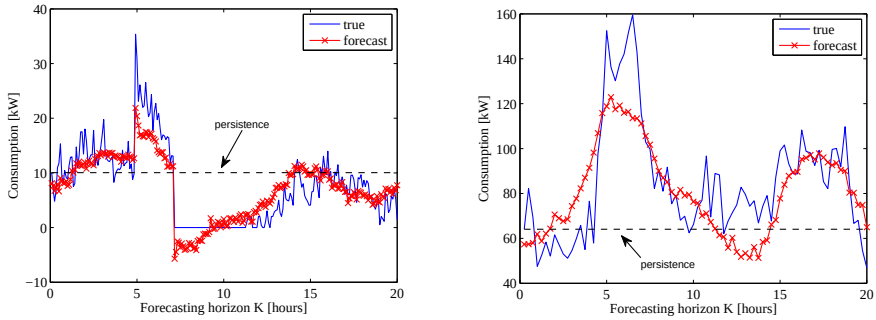
$$\hat{c}_{t+k|t} = \begin{pmatrix} 1 & \mathbf{X}_{t+k}^T \end{pmatrix} \boldsymbol{\theta} \quad \text{(5.5)}$$

as no accumulation of errors is present since the consumption is modelled without any dynamics.

The forecasting performance of the trained models should be evaluated on a *test* set: a dataset on which the model has not be trained on. This mimics the situation of unknown data being presented to the model. At each time point of the training set, and for each prediction step $k = 1, \ldots, K$, the k-step prediction is computed and compared to the real value, yielding a squared prediction error $(\hat{c}_{t+k|t} - c_{t+k})^2$.

Additionally, it is interesting to compare this prediction against the simplest prediction model available: the *persistence*. The latter assumes no knowledge of any pattern or structure of the data, and is expressed simply as $\hat{c}_{t+k|t} = c_t$. In other

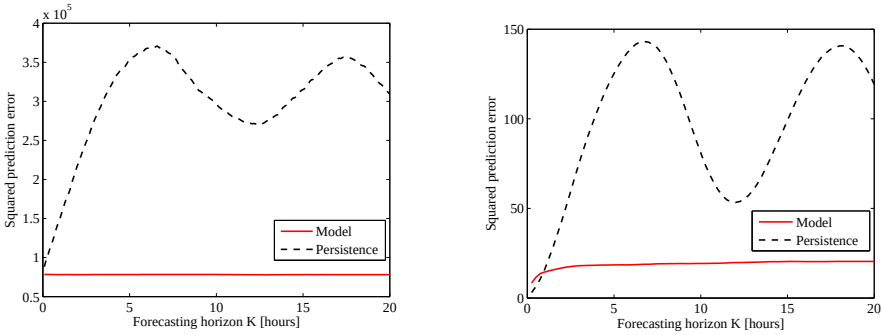
words, the persistence predicts that future values will be the same as the current value.



(a) *Simulated data.* **(b)** *Olympic Peninsula.*

Figure 5.2: *Exemplary forecasts of consumption models and persistence. Note that the simulated consumption is sometimes forecasted to be negative.*

The squared predictions errors are then averaged for each horizon K, and the results are depicted in Figure 5.3 both for the simulated (a) and the real data set (b).



(a) *Simulated data* **(b)** *Olympic Peninsula*

Figure 5.3: *Forecasting performance of models obtained from the simulated (a) and the Olympic Peninsula data set (b). A constant error is observed in the simulated data set, as the uncertainty only depends on external variables. On the Olympic Peninsula data set however, a short accumulation of error is observed. The periodicity of consumption itself yields periodicity in the persistence.*

On the simulated data set, it is found that the forecast error does not grow with the horizon, as expected because there should be no accumulation of errors (the forecast is only driven by the external variables). Compared to the persistence, it seems natural that any knowledge of an underlying system improves the forecasting abilities, as shown in Figure 5.3a where the error related to the persistence is fairly large. A major limitation arises from the fact that the forecasted consumption can drop below zero, as seen in Figure 5.2a. The problem lies in the linearity of the model, which can not approximate well enough the strong non-linearity occurring when the consumption reaches zero.

Regarding the Olympic Peninsula data set, consumption dynamics have to be included as seen in Section 4.3.2. Using the ARX structure presented in Section 4.3.2, the accumulation of prediction errors can not be avoided as seen in Figure 5.3b.

The persistence prediction error is found to fluctuate, reproducing the two peaks of the daily consumption pattern previously studied. This comes from the fact that the morning and evening peaks are similar, yielding a similar prediction error. The persistence prediction error is therefore autocorrelated with a period of approximately 12 hours, the latter being the time between the morning and evening consumption peaks.

An interesting fact about the real data is that the persistence performs slightly better than the model for horizons K smaller that one hour. This leads to the conclusion that the model predicts better the overall long-term mean instead of individual point forecasts.

Nevertheless, one should always keep in mind that in order for the models to be used in a one-way communication implementation, all variables must be measurable from the outside on an aggregated level. The temperature setpoint T_s is not measurable from the outside. Hence, the next section presents an extension estimating this hidden variable.

5.2.3 Forecasting the heating setpoint

Computing the conditional mean of the heating setpoint T_s conditioned on the minute of the day (using the method in Section 4.1) reveals a clear non-linear dependency on the minute of the day (Figure 5.4).

One way to build a model describing the non-linear relation of T_s on the minute

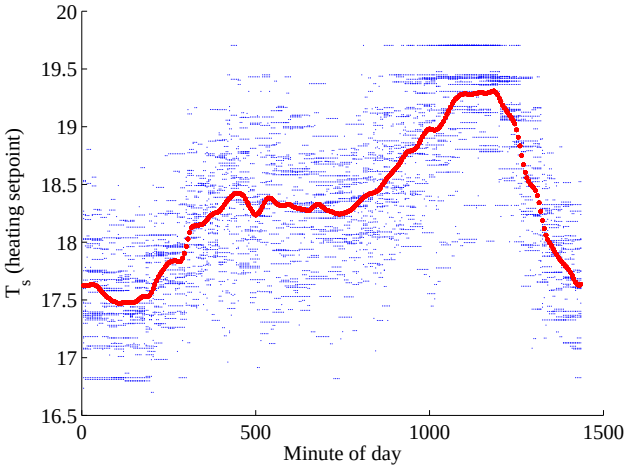


Figure 5.4: *Estimation of the relationship between the minute of day and the heating setpoint in the Olympic Peninsula dataset by using a kernel smoother estimating the conditional mean.*

of the day u is to make use of a basis function $\mathbf{b}(u)^T$ which will be linearly combined to model the setpoint T_s.

Let us introduce the so-called Fourier basis for a given set of N frequencies $f_1, f_2, \ldots, f_N$.

$$\mathbf{b}(u)^T = \begin{pmatrix} 1 & \sin(2\pi f_1 u) & \cos(2\pi f_1 u) & \ldots & \sin(2\pi f_N u) & \cos(2\pi f_N u) \end{pmatrix} \quad (5.6)$$

The unknown parameter vector $\boldsymbol{\theta}$ is introduced such that

$$T_s = \mathbf{b}(u)^T \boldsymbol{\theta} + \epsilon \quad (5.7)$$

Stacking n observations vertically, we obtain

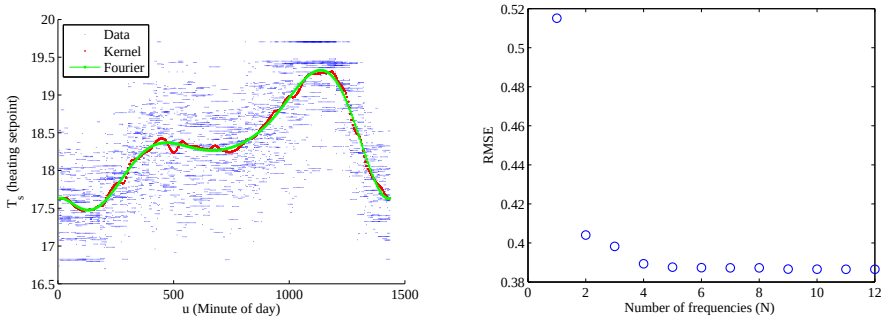
$$\begin{pmatrix} T_{s1} \\ \vdots \\ T_{sn} \end{pmatrix} = \begin{pmatrix} \mathbf{b}(u_1)^T \\ \vdots \\ \mathbf{b}(u_n)^T \end{pmatrix} \boldsymbol{\theta} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (5.8a)$$

$$\mathbf{T_s} = \mathbf{B}^T \boldsymbol{\theta} + \boldsymbol{\epsilon} \quad (5.8b)$$

The parameters $\boldsymbol{\theta}$ can then be estimated by computing the least squares estima-

tor [15]

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{T_s} \tag{5.9}$$



(a) *Fitting the heating setpoint by using a linear model with a Fourier basis, see Equation* (5.8). **(b)** *Root mean square error (RMSE) as a function of the numbers of harmonics used.*

Figure 5.5

The temperature setpoint, due to its daily pattern, has a period of approximately 24 hours. Therefore, the smallest frequency f_1 should correspond to a period of one day. Multiples of this frequency, called harmonics, are then added as extra frequencies. The number of harmonics is chosen by minimizing the root mean square error (RMSE) of the fit (Figure 5.5b). In that sense, $N = 4$ harmonics are found to be sufficient.

5.2.4 Forecasting without knowledge of the heating setpoint with an NFIR

Even though a model for the temperature setpoint T_s has been established in Section 5.2.3, this model has to be trained on data which can be hard to obtain. A model based exclusively on external inputs is then sought by replacing the linear dependency on T_s by a non-linear dependency g on the minute of the day u_t. We will refer to it as the *Non-linear Finite-Impulse-Response model (NFIR)*

Using the FIR model from Section 4.3.2, let us separate the linear inputs $\mathbf{X}_t$ and

the non-linear input u_t (minute of the day)

$$\begin{aligned} c_t &= \begin{pmatrix}1 & \mathbf{X}_t^T\end{pmatrix}\boldsymbol{\theta}_x + \begin{pmatrix}g(u_t) & g(u_{t-1}) & \dots & g(u_{t-n_u-1})\end{pmatrix}\begin{pmatrix}\theta_{u,1}\\ \theta_{u,2}\\ \vdots\\ \theta_{u,n_u}\end{pmatrix} \\ &= \begin{pmatrix}1 & \mathbf{X}_t^T\end{pmatrix}\boldsymbol{\theta}_x + \mathbf{G}^T\boldsymbol{\theta}_u, \end{aligned} \tag{5.10}$$

where the inputs $\mathbf{X}_t$ do not contain the temperature setpoint anymore.

The non-linear relation $g(u)$ can then be approximated by a linear model with a basis function $g(u) \approx \mathbf{b}(u)^T\boldsymbol{\theta}_b$. For example, using a second order polynomial basis would yield $\mathbf{b}(u)^T = \begin{pmatrix}1, u, u^2\end{pmatrix}$ with $\boldsymbol{\theta}_b$ being the polynomial coefficients. Those vectors do not occur in the parameter estimation, and can therefore be computed beforehand. For the purpose of modelling the non-linear dependency of the setpoint on the minute of the day, the Fourier basis function of Equation (5.6) will be used, with the 4 corresponding frequencies found in Section 5.2.3. Rewriting to isolate the basis coefficients $\boldsymbol{\theta}_b$ yields

$$\begin{aligned} c_t &= \begin{pmatrix}1 & \mathbf{X}_t^T\end{pmatrix}\boldsymbol{\theta}_x + \begin{pmatrix}\mathbf{b}(u_t)^T\boldsymbol{\theta}_b & \mathbf{b}(u_{t-1})^T\boldsymbol{\theta}_b & \dots & \mathbf{b}(u_{t-n_u-1})^T\boldsymbol{\theta}_b\end{pmatrix}\begin{pmatrix}\theta_{u,1}\\ \theta_{u,2}\\ \vdots\\ \theta_{u,n_u}\end{pmatrix} \\ &= \begin{pmatrix}1 & \mathbf{X}_t^T\end{pmatrix}\boldsymbol{\theta}_x \\ &\quad + \begin{pmatrix}\mathbf{b}(u_t)^T & \mathbf{b}(u_{t-1})^T & \dots & \mathbf{b}(u_{t-n_u-1})^T\end{pmatrix}\begin{pmatrix}\boldsymbol{\theta}_b & 0 & \dots & 0\\ 0 & \boldsymbol{\theta}_b & \dots & 0\\ \vdots & \vdots & \ddots & \dots\\ 0 & 0 & \dots & \boldsymbol{\theta}_b\end{pmatrix}\begin{pmatrix}\theta_{u,1}\\ \theta_{u,2}\\ \vdots\\ \theta_{u,n_u}\end{pmatrix} \\ &= \begin{pmatrix}1 & \mathbf{X}_t^T\end{pmatrix}\boldsymbol{\theta}_x + \mathbf{B}_t^T\begin{pmatrix}\boldsymbol{\theta}_b & 0 & \dots & 0\\ 0 & \boldsymbol{\theta}_b & \dots & 0\\ \vdots & \vdots & \ddots & \dots\\ 0 & 0 & \dots & \boldsymbol{\theta}_b\end{pmatrix}\begin{pmatrix}\theta_{u,1}\\ \theta_{u,2}\\ \vdots\\ \theta_{u,n_u}\end{pmatrix} \end{aligned} \tag{5.11}$$

By defining the Kronecker product between two matrices A and B, of sizes m-by-

n and p-by-q respectively, as the following mp-by-nq matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}, \tag{5.12}$$

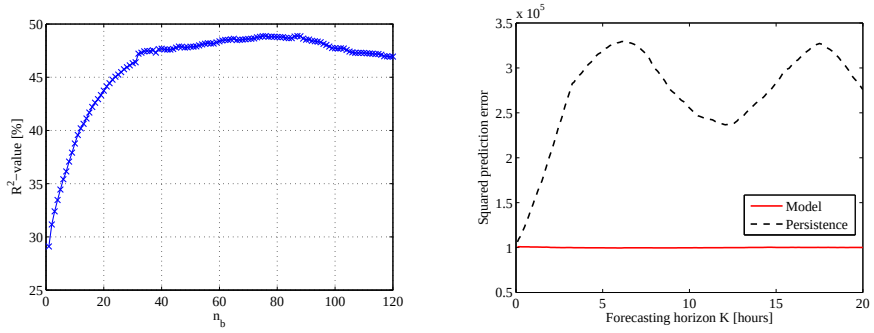
Equation (5.11) can then handily be rewritten as

$$c_t = \begin{pmatrix} 1 & \mathbf{X}_t^T \end{pmatrix} \boldsymbol{\theta}_x + \mathbf{B}_t^T \left(\mathbf{I}_{n_u} \otimes \boldsymbol{\theta}_b \right) \boldsymbol{\theta}_u, \tag{5.13}$$

where $\mathbf{I}_{n_u}$ is the n_u-by-n_u identity matrix. By using the relation $(\mathbf{I}_p \otimes \mathbf{b})\mathbf{a} = \mathbf{a} \otimes \mathbf{b}$ [4], this equation can furthermore by rewritten to the compact notation

$$c_t = \begin{pmatrix} 1 & \mathbf{X}_t^T \end{pmatrix} \boldsymbol{\theta}_x + \mathbf{B}_t^T \left(\boldsymbol{\theta}_u \otimes \boldsymbol{\theta}_b \right) \tag{5.14}$$

The product of the basis coefficients $\boldsymbol{\theta}_b$ with the lag coefficients $\boldsymbol{\theta}_u$ makes this model non-linear in the parameters, with $n_x + n_u + n_b$ coefficients (n_b being the number of components in the basis vector). The uniqueness of the optimal parameters vanishes as several combinations of $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_b$ can provide the fitting performances. One could estimate the parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}_u \otimes \boldsymbol{\theta}_b$, which then makes the model linear in $\boldsymbol{\theta}$. In this case, the number of parameters then grows to $n_x + n_u \cdot n_b$. This is because we ignore the crucial condition that a common $\boldsymbol{\theta}_b$ is used for each input signal u_{t-k} at each lag k, manifesting that the same non-linear function $g(u_{t-k})$ is used for all lags.



(a) $n_b = 40$ *coefficients are found to be necessary, yielding 168 parameters for this model (compared to 321 for the FIR model of Section 4.3.2).* **(b)** *Forecasting performances are very similar to the FIR model of Section 4.3.2.*

In order to find the optimal set of parameters the sum of squared prediction errors is minimised, where the prediction errors (residuals) are formulated as

$$r_t(\boldsymbol{\theta}_x, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) = c_t - \begin{pmatrix} 1 & \mathbf{X}_t^T \end{pmatrix} \boldsymbol{\theta}_x - \mathbf{B}_t^T \left(\boldsymbol{\theta}_u \otimes \boldsymbol{\theta}_b \right). \tag{5.15}$$

Due to the non-linearity in (5.15), we have to solve a non-linear least squares problem which is studied in detail in [16]. For this purpose we make use of the *Levenberg-Marquardt* algorithm which is based on a combination of the Gauss-Newton method and the gradient descend method[1].

Solving performances are increased if the exact Jacobian of the residual function,

$$J(r_t(\boldsymbol{\theta}_x, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b)) = \begin{pmatrix} \nabla_{\boldsymbol{\theta}_x} r_t(\boldsymbol{\theta}_x, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) & \nabla_{\boldsymbol{\theta}_u} r_t(\boldsymbol{\theta}_u, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) & \nabla_{\boldsymbol{\theta}_b} r_t(\boldsymbol{\theta}_b, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) \end{pmatrix} \tag{5.16}$$

is supplied instead of a first order difference approximation. The gradients with respect to each parameter vector are then found to be

$$\nabla_{\boldsymbol{\theta}_x} r_t(\boldsymbol{\theta}_x, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) = -\begin{pmatrix} 1 & \mathbf{X}_t^T \end{pmatrix} \tag{5.17a}$$

$$\nabla_{\boldsymbol{\theta}_u} r_t(\boldsymbol{\theta}_x, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) = -\mathbf{B}_t^T \left(\mathbf{I}_{n_u} \otimes \boldsymbol{\theta}_b\right) \tag{5.17b}$$

$$\nabla_{\boldsymbol{\theta}_b} r_t(\boldsymbol{\theta}_x, \boldsymbol{\theta}_u, \boldsymbol{\theta}_b) = -\left(\boldsymbol{\theta}_u \otimes \mathbf{I}_{nb}\right) \tag{5.17c}$$

Based on the simulated data, the model order and performance assessment of the NFIR is then done in the same manner as in Section 4.3.2. Compared to the FIR model obtained in Section 4.3.2, the performance of the NFIR is fairly satisfactory, as it only reduces the R^2-value from 59.1% to 47.7%. Note however that only half as many parameters are needed.

5.3 Recursive and adaptive estimation

In an online implementation of a forecasting system, one needs to take into account every new measurement available. The more measurements the model is estimated on, the less uncertain the model parameters become. The simplest approach is to re-estimate the model on the whole available dataset every time a new measurement is available. This is called *offline* estimation. However, a *recursive* estimation enables to take into account the previous model estimation and to only correct the latter by the innovation that the new measurement contains, ending up in an *online* estimation.

[1] Implementation taken from DTU-IMM [18] as it yields faster computation times than MATLAB's version

5.3.1 Linear recursive and adaptive estimation

Considering a linear model with gaussian distributed errors ϵ_t

$$y_t = \mathbf{x}_t^T \boldsymbol{\theta} + \epsilon_t,$$

the model parameters $\boldsymbol{\theta}$ can be estimated as the solution of the minimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{t} (y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2 \right\}. \tag{5.18}$$

By vertically stacking observations such that the matrix $\mathbf{X}_t$ and and the output vector $\mathbf{y}_t$ contains observations up to time t, the solution to Equation (5.18) is found to be the least squares estimator [15]

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}_t^T \mathbf{X}_t)^{-1} \mathbf{X}_t^T \mathbf{y_t}. \tag{5.19}$$

The *offline* solution in equation (5.19) can be rewritten to

$$\hat{\boldsymbol{\theta}}_t = \left(\sum_{i=1}^{t} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left(\sum_{i=1}^{t} \mathbf{x}_i y_i \right) = \mathbf{R}_t^{-1} \mathbf{h}_t,$$

where it can be seen that the current parameter estimate $\hat{\boldsymbol{\theta}}_t$ can be expressed in terms of the old estimate and the new observation

$$\hat{\boldsymbol{\theta}}_t \;=\; \mathbf{R}_{t-1}^{-1} \mathbf{h}_{t-1} + \left(\mathbf{x}_t \mathbf{x}_t^T \right)^{-1} \mathbf{x}_t y_t = \hat{\boldsymbol{\theta}}_{t-1} + \left(\mathbf{x}_t \mathbf{x}_t^T \right)^{-1} \mathbf{x}_t y_t$$

Further rearranging (see [15]) leads to the final update equation for the recursive least squares estimation

$$\hat{\boldsymbol{\theta}}_t \;=\; \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{R}_t^{-1} \mathbf{x}_t \left(y_t - \mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{t-1} \right) \quad \text{with} \tag{5.20a}$$

$$\mathbf{R}_t \;=\; \mathbf{R}_{t-1} + \mathbf{x}_t \mathbf{x}_t^T \tag{5.20b}$$

Equations (5.20) above have a strong intuitive meaning. The current parameter estimate $\hat{\boldsymbol{\theta}}_t$ is obtained by adding a correction to the previous estimate $\hat{\boldsymbol{\theta}}_{t-1}$. The correction is proportional to the one-step-ahead prediction error based on the previous parameter estimate $(y_t - \mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{t-1})$.

So far, the offline solution has just been reformulated in a recursive form. However, by introducing a forgetting factor, one can weight down old observations that might be obsolete. In a non stationary system, parameters evolve with time. This is the case if suddenly new appliances are installed in a home, or if a new

house is built. Furthermore, price variations introduce a change in the price reference upon which the standardized price is computed. Because the models are build on the original (observable) price signal, the response parameters changes if the price reference changes. Those temporal changes indicate the need of having an *adaptive* model estimation by 'forgetting' at a certain rate the past observations, and by putting more emphasis on recent observations, more representative of the current system's behaviour. A forgetting factor α is therefore introduced in the minimization problem, putting exponentially decaying weights on past observations:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{t} \alpha^{t-i} (y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2 \right\}. \tag{5.21}$$

Deriving the recursive formulation again leads to a modification of the update formula of equation (5.20b) such that

$$\mathbf{R}_t \quad = \quad \alpha \mathbf{R}_{t-1} + \mathbf{x}_t \mathbf{x}_t^T, \tag{5.22}$$

where $0 < \alpha \leq 1$. Note, that for $\alpha = 1$ we obtain the offline solution from Equation (5.19), see Figure 5.6.

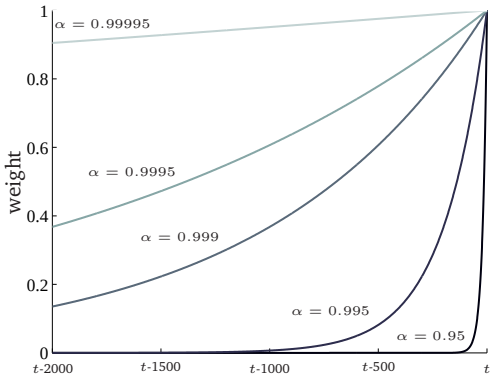


Figure 5.6: *Exponential weighting of past observations for different forgetting factors α and a sample rate of one.*

In order to get an idea of how many samples are considered for the estimation (those that are not removed because they have become obsolete), we introduce the effective number of samples N^* which can be computed as

$$N^* = \frac{1}{1-\alpha}. \tag{5.23}$$

Note that if the inputs $\mathbf{x}$ are not normalized, the matrix $\mathbf{R}_t$ can become singular and inverting it can become difficult. Care has therefore been taken by normalizing the different inputs in the recursive and adaptive implementations.

5.3.2 Non-linear recursive and adaptive estimation

In order to be able to use the NFIR model of Section 5.2.4, one must establish recursive and adaptive estimators for the non-linear case.

The set of parameters $\boldsymbol{\theta}_t$ at time t is found by minimizing the squared prediction errors up to time t, weighted by a forgetting factor α

$$S_t(\boldsymbol{\theta}_t) = \sum_{i=1}^{t} \left\| \alpha^{t-i} \left(y_i - f(\mathbf{x}_i, \boldsymbol{\theta}_t) \right) \right\|^2, \tag{5.24}$$

where f is the non-linear prediction model. The optimal parameter vector $\boldsymbol{\theta}_t$ at time t is found by minimizing $S_t(\boldsymbol{\theta}_t)$ by putting the Jacobian with respect to the parameters to zero,

$$\begin{aligned} 0 &= \nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t) \\ &= -2 \sum_{i=1}^{t} \alpha^{t-i} \nabla_{\boldsymbol{\theta}} f(\mathbf{x}_i, \boldsymbol{\theta}_t) \left(y_i - f(\mathbf{x}_i, \boldsymbol{\theta}_t) \right) \\ &= -2 \sum_{i=1}^{t} \alpha^{t-i} \boldsymbol{\psi}_i(\boldsymbol{\theta}_t) \epsilon_i(\boldsymbol{\theta}_t), \end{aligned} \tag{5.25}$$

where $\boldsymbol{\psi}_i(\boldsymbol{\theta}_t) = \nabla_{\boldsymbol{\theta}} f(\mathbf{x}_i, \boldsymbol{\theta}_t)$ and $\epsilon_i(\boldsymbol{\theta}_t) = y_i - f(\mathbf{x}_i, \boldsymbol{\theta}_t)$.

Furthermore, the Hessian matrix is found as

$$\mathbf{H}_t(\boldsymbol{\theta}_t) = 2 \sum_{i=1}^{t} \alpha^{t-i} \left(\boldsymbol{\psi}_i(\boldsymbol{\theta}_t) \boldsymbol{\psi}_i^T(\boldsymbol{\theta}_t) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\psi}_i(\boldsymbol{\theta}_t) \epsilon_i(\boldsymbol{\theta}_t) \right). \tag{5.26}$$

Note that in a region close to the true minimum $\boldsymbol{\theta}_t^*$, the second term is close to zero. Neglecting that term ensures a positive definite Hessian matrix. Given a non-optimal solution $\boldsymbol{\theta}_t^0$ in a region close to the true minimum, the linearized approximation of $\nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t^*)$ around $\boldsymbol{\theta}_t^0$ can be written as

$$\nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t^*) = \nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t^0) + \mathbf{H}_t(\boldsymbol{\theta}_t^0)(\boldsymbol{\theta}_t^* - \boldsymbol{\theta}_t^0) \tag{5.27}$$

such that an approximation to the optimal parameter $\boldsymbol{\theta}_t^*$ is obtained by putting $\nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t^*) = 0$

$$\boldsymbol{\theta}_t^* = \boldsymbol{\theta}_t^0 - \left[\mathbf{H}_t(\boldsymbol{\theta}_t^0)\right]^{-1} \nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t^0). \tag{5.28}$$

By iteratively applying Equation (5.28), one finds the *offline* solution given an initial parameter guess $\boldsymbol{\theta}_t^0$. This iterative method is called Newton's method, or the Newton-Raphson method. Note that if $S_t(\boldsymbol{\theta}_t)$ is quadratic in $\boldsymbol{\theta}_t$ such that $\nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t)$ is linear in $\boldsymbol{\theta}_t$, the optimal parameter $\boldsymbol{\theta}_t^*$ is obtained with one iteration.

Based on Equation (5.25), the Jacobian can be recursively expressed. Assuming optimality of the previous parameter $\boldsymbol{\theta}_{t-1}^*$ such that $\nabla_{\boldsymbol{\theta}} S_{t-1}(\boldsymbol{\theta}_{t-1}) = 0$,

$$\nabla_{\boldsymbol{\theta}} S_t(\boldsymbol{\theta}_t) = \nabla_{\boldsymbol{\theta}} S_{t-1}(\boldsymbol{\theta}_{t-1}) - 2\boldsymbol{\psi}_t(\boldsymbol{\theta}_t)\epsilon_t(\boldsymbol{\theta}_t) = -2\boldsymbol{\psi}_t(\boldsymbol{\theta}_t)\epsilon_t(\boldsymbol{\theta}_t). \tag{5.29}$$

Furthermore assuming that the previous parameter is close to the optimal solution at the next time index t, the Hessian matrix can be rewritten in the following recursive manner

$$\begin{aligned} \mathbf{H}_t(\boldsymbol{\theta}_t) &= 2\sum_{i=1}^{t} \alpha^{t-i} \boldsymbol{\psi}_i(\boldsymbol{\theta}_t)\boldsymbol{\psi}_i^T(\boldsymbol{\theta}_t) \\ &= \alpha \mathbf{H}_{t-1}(\boldsymbol{\theta}_{t-1}) + 2\boldsymbol{\psi}_t(\boldsymbol{\theta}_t)\boldsymbol{\psi}_t^T(\boldsymbol{\theta}_t). \end{aligned} \tag{5.30}$$

Using the notation $\mathbf{R}_t = \frac{1}{2}\mathbf{H}_t$, the recursive prediction error method [15] can be established by combining Equations (5.28), (5.29) and (5.30):

$$\mathbf{R}_t = \alpha \mathbf{R}_{t-1} + \boldsymbol{\psi}_t \boldsymbol{\psi}_t^T \tag{5.31a}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mathbf{R}_t^{-1} \boldsymbol{\psi}_t \epsilon_t \tag{5.31b}$$

5.4 Controlling the consumption by price

5.4.1 Generalized Minimum Variance (GMV) controller

The aim of the controller is to generate a price for the next sample $p_{t,1}$ by minimizing the future expected deviations from a consumption reference and by penalizing the control signal (the price), as Equation (1.4) states it in the problem formulation in Section 1.3. The hard constraints imposed on the price and consumption will be ignored for now, and we will assume a one-step consumption

penalty $w_{t,1} = 1$. As derived in the problem formulation, the loss function to be minimized is

$$\begin{aligned} L(p_{t,1}) &= \mathbb{E}\left\{ \left\| c_{t,1}\left(p_{t,1}, \mathcal{F}_t\right) - c^*_{t,1} \right\|^2 + \lambda_{t,1} \| p_{t,1} - p^*_{t,1} \|^2 \middle| \mathcal{F}_t \right\} \\ &= \left(\mathbb{E}\left\{ c_{t,1}\left(p_{t,1}, \mathcal{F}_t\right) | \mathcal{F}_t \right\} - c^*_{t,1} \right)^2 \\ &\quad + \mathrm{Var}\left\{ c_{t,k}\left(p_{t,1}, \mathcal{F}_t\right) | \mathcal{F}_t \right\} \\ &\quad + \lambda_{t,1} \left(p_{t,1} - p^*_{t,1} \right)^2 \end{aligned} \tag{5.32}$$

where the conditional expectation $\mathbb{E}\left\{ c_{t,1}\left(p_{t,1}, \mathcal{F}_t\right) | \mathcal{F}_t \right\}$ is the optimal one-step predictor of the consumption, and the conditional variance $\mathrm{Var}\left\{ c_{t,1}\left(p_{t,1}, \mathcal{F}_t\right) | \mathcal{F}_t \right\}$ is the associated uncertainty.

Prediction using the FIR model found in Section 4.3.2 can be done according to Section 5.2.1. By furthermore separating controllable (future) price $p_{t,1}$ and uncontrollable (past) prices $\mathbf{p}_t$,

$$\begin{aligned} \mathbb{E}\left\{ c_{t,1}\left(p_{t,1}, \mathcal{F}_t\right) | \mathcal{F}_t \right\} = \begin{pmatrix} 1 & \mathbf{X}^T_{t+1} \end{pmatrix} \boldsymbol{\theta} &= \begin{pmatrix} p_{t,1} & 1 & \mathbf{p}_t^T & \hat{\mathbf{Z}}^T_{t+1} \end{pmatrix} \begin{pmatrix} \theta_p \\ \theta_0 \\ \boldsymbol{\theta}_{\mathbf{p}} \\ \boldsymbol{\theta}_{\hat{Z}} \end{pmatrix} \\ &= p_{t,1}\theta_p + \mathbf{Y}_t^T \boldsymbol{\theta}_Y, \end{aligned} \tag{5.33}$$

where the matrix $\mathbf{Y}_t$ contains an intercept, prices up to time t and other external variables up to time $t+1$. It has also been assumed that the model errors are i.i.d. Gaussian with zero mean.

Similarly, prediction and decoupling for the NFIR model found in Section 5.2.4 is achieved by

$$\begin{aligned} \mathbb{E}\left\{ c_{t,1}\left(p_{t,1}, \mathcal{F}_t\right) | \mathcal{F}_t \right\} &= \begin{pmatrix} 1 & \mathbf{X}^T_{t+1} \end{pmatrix} \boldsymbol{\theta}_x + \mathbf{B}^T_{t+1} \left(\boldsymbol{\theta}_u \otimes \boldsymbol{\theta}_b \right) \\ &= \begin{pmatrix} p_{t,1} & 1 & \mathbf{p}_t^T & \hat{\mathbf{Z}}^T_{t+1} & \mathbf{B}^T_{t+1} \end{pmatrix} \begin{pmatrix} \theta_{x,p} \\ \theta_{x,0} \\ \boldsymbol{\theta}_{x,\mathbf{p}} \\ \boldsymbol{\theta}_{x,\hat{Z}} \\ \boldsymbol{\theta}_u \otimes \boldsymbol{\theta}_b \end{pmatrix} \\ &= p_{t,1}\theta_p + \mathbf{Y}^T_{t+1} \boldsymbol{\theta}_Y. \end{aligned} \tag{5.34}$$

Note that even though the NFIR is non-linear in its parameters, it is linear in the control signal $p_{t,1}$. Furthermore, in both the FIR and the NFIR, $\mathbf{Y}_t^T \boldsymbol{\theta}_Y$ can easily

be found by setting the future price to zero, forming the equality

$$\mathbb{E}\left\{c_{t,1}\left(0,\mathcal{F}_t\right)|\mathcal{F}_t\right\} = \mathbf{Y}_{t+1}^T\boldsymbol{\theta}_Y. \tag{5.35}$$

Finding the minimum of the loss function is done by setting the derivative to zero and by checking that the second derivative is positive.

$$\begin{aligned}\frac{\partial}{\partial p_{t,1}}L(p_{t,1}) = 2\theta_p\left(p_{t,1}\theta_p + \mathbf{Y}_{t+1}^T\boldsymbol{\theta}_y - c_{t,1}^*\right) &+ \frac{\partial}{\partial p_{t,1}}\mathrm{Var}\left\{c_{t,k}\left(p_{t,1},\mathcal{F}_t\right)|\mathcal{F}_t\right\} \\ &+ 2\lambda_{t,1}\left(p_{t,1} - p_{t,1}^*\right)\end{aligned} \tag{5.36}$$

Supposing that the uncertainty on the predictor does not depend on $p_{t,1}$, one can isolate $p_{t,1}$

$$\begin{aligned}0 &= 2p_{t,1}\left(\theta_p^2 + \lambda_{t,1}\right) + 2\theta_p\left(\mathbf{Y}_{t+1}^T\boldsymbol{\theta}_y - c_{t,1}^*\right) - 2\lambda_{t,1}p_{t,1}^* \\ p_{t,1} &= \frac{\theta_p\left(c_{t,1}^* - \mathbf{Y}_{t+1}^T\boldsymbol{\theta}_y\right) + \lambda_{t,1}p_{t,1}^*}{\theta_p^2 + \lambda_{t,1}}.\end{aligned} \tag{5.37}$$

By calculating the second derivative, we indeed verify that the solution is a minimum:

$$\frac{\partial^2}{\partial p_{t,1}^2}L(p_{t,1}) = 2\left(\theta_p^2 + \lambda_{t,1}\right) \geq 0. \tag{5.38}$$

One should note that by introducing the control law of Equation (5.37), the one-step ahead price was found as the price yielding the reference $c_{t,1}^*$ as optimal one-step prediction if $\lambda_{t,1} = 0$. The control error is then exactly the prediction error against which the model is recursively trained. In an adaptive model estimation, the parameters will be then chosen such that the *control* error is minimized: the system is therefore self-tuning.

Without a price penalty, the price signal is unstable (Figure 5.7) due to an unreachable consumption reference. Situations of unreachable consumption references occur e. g. during the night, where there is not enough heating needs to increase the consumption to the desired level. Note that even if the consumption reference is chosen close to the mean consumption as in Figure 5.7, the price signal becomes unstable after a couple of days.

Price penalties up to $\lambda_{t,1} = 10^7$ were tested, but none could ensure stability in every situation. This could arise from the fact that the controller needs to be able to look more than one step ahead.

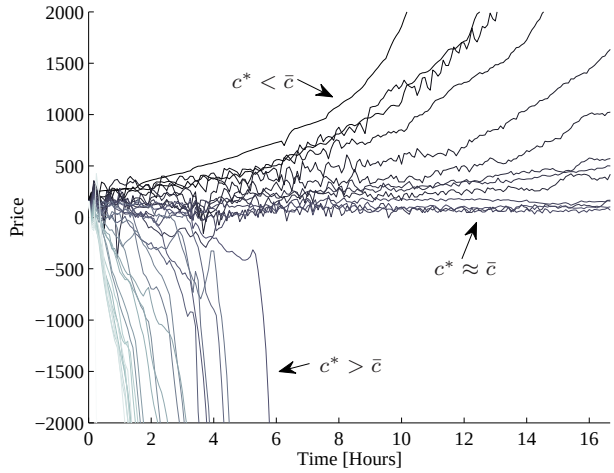


Figure 5.7: *Without a price penalty ($\lambda = 0$), the price signal is unstable. Its divergence speed and direction however depends on the consumption reference c^*, which is here chosen proportional to the mean consumption $\bar{c}$ of the training set. The divergence of the price can be seen as a limitation of a linear controller, which is going to push the prices to extremes in order to follow a (possibly unreachable) consumption target.*

5.4.2 Generalized Predictive Controller (GPC)

Referring to Equation (1.4) and (1.7) from the problem formulation of Section 1.3, the K-step control problem is formulated as

$$
\begin{aligned}
L(p_{t,1},\dots,p_{t,K}) &= \mathbb{E}\Bigg\{ \sum_{k=1}^{K} w_{t,k} \left\| c_{t,k}\left(p_{t,1},\dots,p_{t,k},\mathcal{F}_t\right) - c^*_{t,k} \right\|^2 \\
&\qquad + \lambda_{t,k} \left\| p_{t,k} - p^*_{t,k} \right\|^2 \Bigg| \mathcal{F}_t \Bigg\} \\
&= \sum_{k=1}^{K} \Bigg(w_{t,k} \Big(\mathbb{E}\left\{ c_{t,k}\left(p_{t,1},\dots,p_{t,k},\mathcal{F}_t\right) | \mathcal{F}_t \right\} - c^*_{t,k} \Big)^2 \\
&\qquad + w_{t,k} \Big(\operatorname{Var}\left\{ c_{t,k}\left(p_{t,1},\dots,p_{t,k},\mathcal{F}_t\right) | \mathcal{F}_t \right\} \Big) \\
&\qquad + \lambda_{t,k} \left(p_{t,k} - p^*_{t,k} \right)^2 \Bigg).
\end{aligned}
$$

If the uncertainty of the predictor is independent of the control signal, then the variance term vanishes in the minimization procedure.

Let us group the future prices $p_{t,k}$, the desired price levels $p^*_{t,k}$ and the desired consumptions $c^*_{t,k}$ up to the horizon K into vectors $\mathbf{p}_t$, $\mathbf{p}^*_t$ and $\mathbf{c}^*_t$. Let us furthermore define the vector $\hat{\mathbf{c}}_t$ as the vector having as k-th element the k-step consumption prediction, and the diagonal matrices $\mathbf{W}_t$ and $\boldsymbol{\lambda}_t$ respectively having as k-th diagonal element the consumption weighting factor $w_{t,k}$ and the price penalty $\lambda^*_{t,k}$. The optimization problem can then be rewritten into the quadratic form

$$L(\mathbf{p}_t) = (\hat{\mathbf{c}}_t - \mathbf{c}^*_t)^T \mathbf{W}_t (\hat{\mathbf{c}}_t - \mathbf{c}^*_t) + (\mathbf{p}_t - \mathbf{p}^*_t)^T \boldsymbol{\lambda}_t (\mathbf{p}_t - \mathbf{p}^*_t). \tag{5.39}$$

For the linear FIR and NFIR models found in Section 4.3.2 and 5.2.4, the k-step prediction can be expressed as a sum of contributions from the control variable $\mathbf{p}_t$ and uncontrollable variables $\mathbf{Y}_{t+k}$ (including forecasted external variables). The contribution of the uncontrollable variables are found by taking the k-step predictor and putting the future price values to zero, removing the influence of future prices.

$$\begin{aligned}
\mathbb{E}\left\{c_{t,k}\left(p_{t,1},\ldots,p_{t,k},\mathcal{F}_t\right)|\mathcal{F}_t\right\} &= \begin{pmatrix}\theta_1 & \ldots & \theta_k\end{pmatrix}\begin{pmatrix}p_{t,k}\\ \vdots \\ p_{t,1}\end{pmatrix} + \mathbb{E}\left\{c_{t,k}\left(0,\ldots,0,\mathcal{F}_t\right)|\mathcal{F}_t\right\} \\
&= \begin{pmatrix}\theta_k & \ldots & \theta_1\end{pmatrix}\begin{pmatrix}p_{t,1}\\ \vdots \\ p_{t,k}\end{pmatrix} + \boldsymbol{\theta}_Y^T \mathbf{Y}_{t+k}. \qquad (5.40)
\end{aligned}$$

where θ_k is the k-th coefficients of the price impulse response. Note that the k-step ahead price is multiplied by the *first* coefficient of the price impulse response. Vertically concatenating the predictions for each step k, we obtain the prediction vector $\hat{\mathbf{c}}_t$ at time t

$$\begin{aligned}
\hat{\mathbf{c}}_t &= \begin{pmatrix}
\theta_1 & 0 & \ldots & 0 & 0 \\
\theta_2 & \theta_1 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\theta_{K-1} & \theta_{K-2} & \ldots & \theta_1 & 0 \\
\theta_K & \theta_{K-1} & \ldots & \theta_2 & \theta_1
\end{pmatrix}\begin{pmatrix}p_{t,1}\\ p_{t,2}\\ \vdots \\ p_{t,K}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\theta}_Y^T \mathbf{Y}_{t+1}\\ \vdots \\ \boldsymbol{\theta}_Y^T \mathbf{Y}_{t+K}\end{pmatrix} \qquad (5.41) \\
&= \boldsymbol{\Pi}\mathbf{p}_t + \mathbf{Z}_t \qquad (5.42)
\end{aligned}$$

where $\mathbf{\Pi}$ is a $K \times K$ lower triangular Toeplitz matrix. The quadratic form can then be written as

$$L(\mathbf{p}_t) = (\mathbf{\Pi}\mathbf{p}_t + \mathbf{Z}_t - \mathbf{c}_t^*)^T \mathbf{W}_t (\mathbf{\Pi}\mathbf{p}_t + \mathbf{Z}_t - \mathbf{c}_t^*) + (\mathbf{p}_t - \mathbf{p}_t^*)^T \boldsymbol{\lambda}_t (\mathbf{p}_t - \mathbf{p}_t^*). \quad (5.43)$$

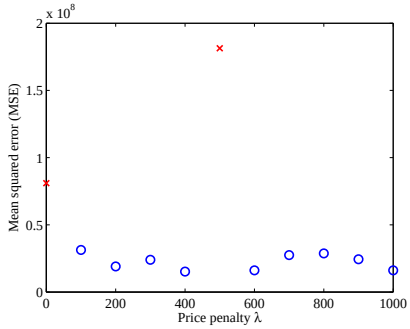
The minimum of this quadratic form is the found by setting the derivative with respect to the parameter vector $\mathbf{p}_t$ to zero.

$$\begin{aligned}
\frac{1}{2}\frac{\partial L(\mathbf{p}_t)}{\partial \mathbf{p}_t} &= (\mathbf{\Pi}\mathbf{p}_t + \mathbf{Z}_t - \mathbf{c}_t^*)^T (\mathbf{W}_t^T \mathbf{\Pi}) + (\mathbf{p}_t - \mathbf{p}_t^*)^T \boldsymbol{\lambda}_t^T && (5.44a)\\
0 &= (\mathbf{\Pi}^T \mathbf{W}_t)(\mathbf{\Pi}\mathbf{p}_t + \mathbf{Z}_t - \mathbf{c}_t^*) + \boldsymbol{\lambda}_t(\mathbf{p}_t - \mathbf{p}_t^*) && (5.44b)\\
0 &= (\mathbf{\Pi}^T \mathbf{W}_t \mathbf{\Pi} + \boldsymbol{\lambda}_t)\mathbf{p}_t + \mathbf{\Pi}^T \mathbf{W}_t(\mathbf{Z}_t - \mathbf{c}_t^*) - \boldsymbol{\lambda}_t \mathbf{p}_t^* && (5.44c)\\
\mathbf{p}_t &= \left[\mathbf{\Pi}^T \mathbf{W}_t \mathbf{\Pi} + \boldsymbol{\lambda}_t\right]^{-1} \left(\boldsymbol{\lambda}_t \mathbf{p}_t^* + \mathbf{\Pi}^T \mathbf{W}_t(\mathbf{c}_t^* - \mathbf{Z}_t)\right) && (5.44d)
\end{aligned}$$

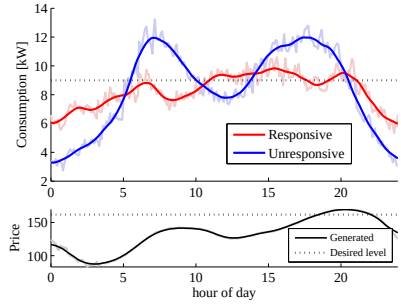
The GPC controller is tested by following a constant consumption target. Two month of consumption data (November and December 2008) are then simulated in order to train the prediction models sufficiently. Starting in January 2009, two month of controlled consumption are simulated (Figure 5.9), and a non-responsive reference group is additionally generated for comparative purposes.

The consumption target c^* is chosen as the mean of the training data, the price level p^* is fixed as the mean of the training prices and the price penalties $\lambda_{t,k}$ are kept constant during the whole horizon. The choice of the prediction horizon K used in the controller is set to the length of the step response $n_b = 40$ ($\approx$ 3.5 hours), as a control action does not have any influence beyond 3.5 hours.

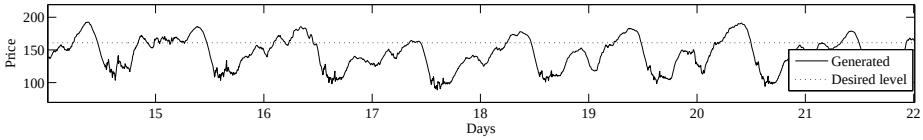
The criterion used to assess the controller is the the mean squared control error, defined as the squared difference between the measured and targeted consumption. For several price penalties λ, the controller is now stable (Figure 5.8a). For $\lambda = 600$, a reduction in peak consumption of nearly 5% is observed together with a mean daily consumption shift of 11%. The price-responsive and unresponsive groups are comparable because controlling by price only decreased the overall consumption by 1%.



(a) *Mean squared control error for different price penalties λ. Red crosses indicate instability, i. e. the price diverges.*



(b) *Performance of the GPC for a constant consumption target using a price penalty of $\lambda = 600$.*



(c) *Daily patterns are observed in the generated price signal.*

Figure 5.8: *The proof-of-concept is illustrated by trying to following a constant reference, yielding a reduction in peak consumption of nearly 5%, a mean daily consumption shift of 11%. The price-responsive and unresponsive groups are comparable as controlling by price only decreased the overall consumption by 1%.*

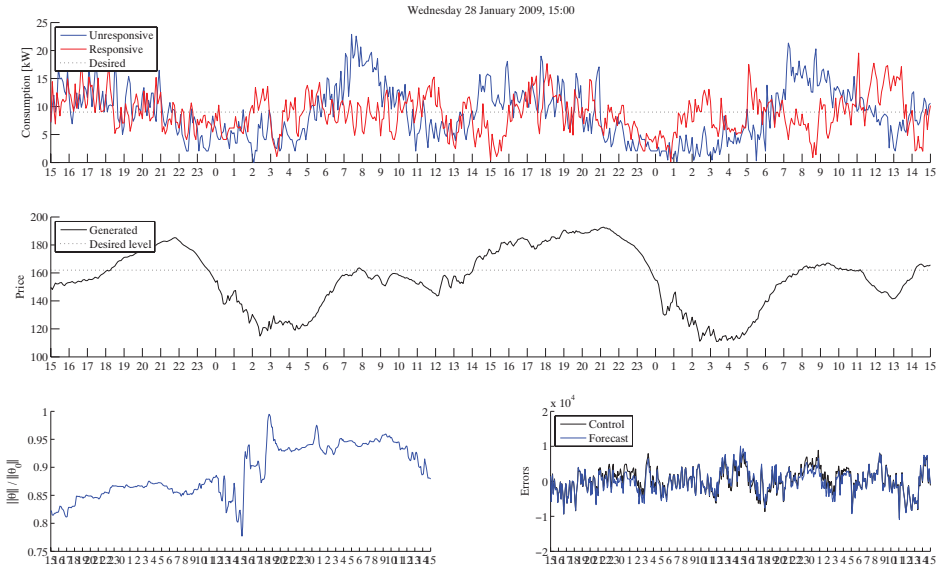


Figure 5.9: *Fast computation times of the simulation framework enabled a real-time computation and visualization of the controlled system.*

Chapter 6

Discussion

Several assumptions have been made when designing and building the different elements of this thesis. The objective of this chapter is to provide a short overview of the limitations we have identified, and the possible improvements possible.

6.1 Further developments

6.1.1 Simulation

- Simple building and appliance models were used. More complex models could be used to better reflect the real dynamics of appliances and buildings.
- All household are affected by the same external variables, like e. g. the outside temperature. Random perturbations could be applied in a stochastic simulation, e. g. reflecting short and local changes in temperature.
- In our simulations, external weather variables are forecasts, not real measurements. Real data should be incorporated, on the same sampling time as the simulation itself.
- Simulation of punctual events (vacation periods, Christmas, etc...) could be included.

6.1.2 Modelling

- For simplicity, our approach only focussed on heating systems. Other devices should be included.
- A more diverse population, possibly clustered, with time-varying price sensitivities and comfort boundaries could be investigated.
- Non-linearities in the system such as e. g. the saturation of price responsivity can possibly better be captured by more advanced models.
- Moving average terms for model errors were not considered. Including them might improve the prediction performance by accounting for correlations of the residuals.
- Punctual events (vacation periods, Christmas time, etc...) could be included into modelling.
- Forecasted values of outside temperature and sun irradiance for the horizon of interest were assumed to be known. How the uncertainty of these forecasted value affect the whole system should be investigated.
- Constraints on the consumption have not been incorporated. Performances are expected to increase by forcing it to be positive.

6.1.3 Control

- Investigate the effect of the time constant involved in the computation of the reference price. It is expected that the stationarity of the price response highly depends on it. As a consequence, it might be easier to achieve stability of the whole control system.
- Investigate the uncertainties of the predictions. By taking advantage of this information, stochastic control can be applied, probably increasing performances.

6.1.4 Concept of controlling by price

- Electricity price control requires a close to real time metering of the consumption, which is not established yet. This is a strong limitation, and further development could include methods to use such a system even though measurements are not available on a small time scale.
- Price reference $\bar{p}$ is not allowed to reach zero, causing an singularity. The time constant involved in its calculation forces the control signal to be unstable if the consumption reference is unreachable.
- Criteria on the control signal must be clearly established due to its variability and high fluctuations.
- Customer incentives need to be added, as this approach mostly look at minimizing the costs of the Balance Responsible. A holistic approach would consider customer costs and benefits into the cost function itself.

6.2 Conclusion

Having the goal of identifying price-responsive heating systems, two data sets were prepared. Understanding the Olympic Peninsula experiment while building a simulation framework provided valuable insights into the dynamics of such a system. The price-responsive behaviour of those two data sets was extracted and modelled, and its dependencies on several other variables were investigated. In the light of a real-life implementation, a non-linear forecasting model was developed based on variables observable on an aggregated level. This model only requires an aggregate metering of household consumptions, together with a price signal and weather forecasts. An adaptive estimation of the forecasting model was then implemented, permitting the development of a price generator in the form of a predictive controller. The proof-of-concept is illustrated by following a constant reference, yielding a reduction in peak consumption of nearly 5%, a mean daily consumption shift of 11%. The price-responsive and unresponsive groups are comparable as controlling by price only decreased the overall consumption by 1%.

Taking a customer perspective is however also necessary. Price signals must be constrained such that customers are guaranteed not to increase their consumption costs. The adoption of such a system will strongly depend on the value created for its user. It can be in form of comfort, as intelligent appliances are installed, or it can be in form of savings, as energy is used during the cheapest hours. Implementing intelligence can also save costs. Taking a heat pump as example, its lifespan is determined by its number of duty cycles. By taking this into account when switching occupancy modes, the device's lifespan could be increased.

There is great potential in such an approach. In Germany, the energy service provider E.ON conducted a study showing that customers can save up to 25% with variable electricity rates[1], concluding that „that electricity consumers change their behaviour if prices are dynamic and their electricity costs and consumption are transparent". The approach is also easily generalizable, as it could work with very different appliances as long as the aggregate population exhibits a price-responsive pattern.

As the tools necessary to simulate and control a price-responsive population have been developed, a natural step is to investigate how much fluctuating energy can be absorbed by a control by price concept. We look forward to participating in such an experiment.

[1] `http://www.eon.com/en/media/news-detail.jsp?id=10432`

Appendix A

Olympic Peninsula dataset description

Variablename	Description
$C(t)$	Total metered consumption at time t μ mean σ^2 variance min minimum value max minimum value N number of metered households
$T_a(t)$	Indoor air temperature measured by the thermostats at time t μ mean σ^2 variance min minimum value max minimum value
$W_i(t)$	Weather of the i'th region (for $i = 1, 2, 3$) at time t Outside temperature Humidity Dewpoint Wind velocity Wind speed Barometer
$S(t)$	Solar irradiance at time t Global horizontal irradiance Direct normal irradiance Diffuse horizontal irradiance
$p(t)$	Price signal at time t
$N_i(t)$	The number of devices among all households active in a certain occupancy mode $\mathcal{O} = i$ at time t
$N_{\text{all}}(t)$	The total number of measured devices among all households at time t
$T_s(\mathcal{O} = i, n)$	Initial heating/cooling setpoint of the n'th device given a certain occupancy mode $\mathcal{O} = i$
$T_s^{\max}(\mathcal{O} = i, n)$	Initial upper heating/cooling setpoint boundary of the n'th device given a certain occupancy mode $\mathcal{O} = i$
$T_s^{\min}(\mathcal{O} = i, n)$	Initial lower heating/cooling setpoint boundary of the n'th device given a certain occupancy mode $\mathcal{O} = i$
$k(\mathcal{O} = i, n)$	Initial price sensitivity factor of the n'th device given a certain occupancy mode $\mathcal{O} = i$
Time	Different time variables indicating Project day $\{1, \ldots, 365\}$ Season of the year $\{1, 2, 3\}$ Month $\{1, \ldots, 12\}$ Weekday $\{1, \ldots, 7\}$ Hour of day $\{0, 1, \ldots, 23\}$ Minute of hour $\{0, \ldots, 55\}$ Weekend/-day Boolean $\{0, 1\}$

Appendix B

Variable selection results

Olympic Peninsula, external variables

Kernel smoothing	Neural network	mRMR with mutual information
MinuteOfDay (66.49944)	MinuteOfDay (61.60608)	Barometer 6.5 (0.10298)
Outside dewpoint 1 h (77.86511)	Outside dewpoint 45 min (76.38716)	IsWeekend (-0.08494)
Barometer 11 h (80.19193)	Outside humidity 8.75 h (77.98602)	Outside temperature 2.75 h (-0.14528)
Outside temperature 10.5 h (83.25824)	IsWeekend (79.37180)	Outside wind speed 10.75 h (-0.16961)
Outside wind speed 10.5 h (86.42221)	Sun: Direct 1 h (81.39006)	Outside wind speed 1.75 h (-0.18757)
Weekday (89.42299)	Barometer 30 min (82.69497)	Outside temperature 10.25 h (-0.20632)
Outside wind speed 0 h (91.97766)	Outside wind direction 1.5 h (83.67840)	Weekday (-0.23392)
Outside wind speed 2 h (93.70091)	Outside wind direction 12 h (84.34848)	Outside temperature 0 h (-0.22298)
Outside wind speed 6.75 h (94.66543)	Outside wind direction 5.5 h (84.63392)	Outside wind speed 9 h (-0.21509)
Outside wind speed 9.75 h (95.47686)	Sun: Global 9.5 h (85.12806)	Outside wind speed 0 h (-0.22336)
Outside wind speed 5.25 h (96.14385)	Outside wind speed 7.25 h (85.28678)	Outside temperature 4.5 h (-0.21968)
Price 3.25 h (96.75166)	Outside wind speed 2.5 h (86.12760)	Outside humidity 9.5 h (-0.22419)
Outside wind speed 11.5 h (97.07072)	Outside wind direction 10.75 h (86.39113)	Outside wind speed 12 h (-0.22348)
Outside wind speed 1.25 h (97.47624)	Outside temperature 10.75 h (86.52029)	Outside wind speed 4 h (-0.23221)
Outside humidity 3 h (97.81783)	Outside humidity 1 h (86.86599)	Outside humidity 11.5 h (-0.24142)

Simulated data, external variables

Kernel smoothing	Neural network	mRMR with mutual information
P 0 h (13.33202)	P 0 (11.16996)	P 0 min (0.11647)
P 3.1 h (24.68210)	P 40 (23.98975)	IsWeekend (-0.11546)
MinuteOfDay 0 (30.56514)	MinuteOfDay (30.95266)	P 5.7 h (-0.14803)
Outside temperature 7.6 h (40.19723)	P 23 (35.77871)	P 11.3 h (-0.20983)
Outside temperature 5.25 h (43.93930)	Outside temperature 0 (37.25690)	P 3.1 h (-0.32849)
P 6.9 h (45.94170)	P 76 (39.90526)	P 8.6 h (-0.33723)
P 5.1 h (47.52105)	P 57 (41.13812)	P 12 h (-0.43432)
P 12 h (48.86143)	Outside temperature 11.8 h (41.55720)	P 10 min (-0.43144)
P 6.7 h (49.88636)	Outside temperature 3.1 h (43.01097)	P 6.8 h (-0.42298)
P 1.6 h (50.59214)	Sun irrandiance 3.4 h (43.46239)	P 4.2 h (-0.43707)
P 9.25 h (51.22167)	Sun irrandiance 5.25 h (44.59067)	P 9.9 h (-0.45363)
IsWeekend 0 h (52.16801)	P 55 min (44.89782)	P 2 h (-0.47854)
Sun irrandiance 0 h (52.77657)	Sun irrandiance 4.7 h (45.49637)	P 7.6 h (-0.49949)
Sun irrandiance 0.8 h (55.56907)	P 11.75 h (45.81029)	P 5 min (-0.48966)
Sun irrandiance 2 h (57.54853)	P 5 min (46.67148)	P 12 h (-0.48774)

Olympic Peninsula, internal and external variables

Kernel smoothing	Neural network	mRMR with mutual information
MinuteOfDay (63.90224)	MinuteOfDay (63.03721)	Heating setpoint 0 h (0.29747)
Outside dewpoint 1 h (77.86511)	Outside dewpoint 1.25 (76.42398)	Inside air temperature 4 h (-0.00092)
Heating setpoint 0 h (80.94186)	Heating setpoint 15 min (78.97656)	IsWeekend 0 h (-0.06443)
Standardized price 11 h (84.11597)	Inside air temperature 2 h (81.03345)	Heating setpoint deviation 5 h (-0.01585)
Outside dewpoint 11.5 h (87.32944)	Outside humidity 10 h (81.58827)	Heating setpoint 6.5 h (-0.06235)
Heating setpoint 7 h (90.10851)	Heating setpoint 9.75 h (82.04817)	Inside air temperature 7.75 h (-0.09483)
Barometer 10.25 h (92.51514)	Outside humidity 1.5 h (83.11926)	Heating setpoint 12 h (-0.11569)
Heating setpoint 12 h (94.22898)	Standardized price 1.75 h (83.85738)	Outside wind speed 9 h (-0.11796)
Outside wind speed 6 h (95.59356)	Barometer 9 h (84.68307)	Heating setpoint 1.25 h (-0.11670)
Heating setpoint deviation 2.75 h (96.76195)	Inside air temperature 7.75 h (85.01700)	Inside air temperature 12 h (-0.13609)
Heating setpoint 2 h (97.49565)	Outside wind speed 12 h (85.48171)	Inside air temperature 1.75 h (-0.13129)
Heating setpoint 9.5 h (98.01686)	Inside air temperature 12 h (86.07103)	Outside temperature 2.75 h (-0.12923)
Outside wind speed 0.5 h (98.41261)	Sun: Direct 1 h (86.54755)	Heating setpoint 7.5 h (-0.12235)
Heating setpoint 10.75 h (98.69975)	Outside wind speed lag 6.25 h (87.38055)	Outside wind speed 1.75 h (-0.13954)
Heating setpoint 8 h (98.92606)	IsWeekend (87.97435)	Inside air temperature 5.75 h (-0.12834)

Simulated data, internal and external variables

Kernel smoothing	Neural network	mRMR with mutual information
Inside air temperature 0 h (46.23483)	Inside air temperature 0 h (45.71687)	Inside air temperature 0 h (0.26025)
Inside air temperature 5 min (67.69149)	Inside air temperature 5 min (72.27457)	IsWeekend (-0.01865)
Inside air temperature 2.8 h (69.94605)	Inside air temperature 3.2 h (80.34203)	Inside air temperature 5.2 h (-0.01064)
Standardized price 5 min (73.37187)	Inside air temperature 1.5 h (82.13722)	Standardized price 0 h (-0.01426)
MinuteOfDay 0 h (78.14871)	Inside air temperature 5.4 h (83.06236)	Inside air temperature 9.1 h (-0.07350)
Inside air temperature 10 min (84.22389)	Outside temperature 5 h (84.04635)	Heating setpoint 0 h (-0.06076)
Inside air temperature 45 min (88.82261)	Inside air temperature 8.8 h (84.40780)	Inside air temperature 3 h (-0.09776)
Inside air temperature 7.4 h (92.15322)	P 2.5 h (84.65704)	Inside air temperature 11.9 h (-0.10374)
Inside air temperature 30 min (94.66637)	Sun irrandiance 55 min (84.93181)	Heating setpoint 6.3 h (-0.10141)
Inside air temperature 15 min (96.21944)	Outside temperature 2.5 h (85.21774)	Inside air temperature 15 min (-0.10966)

Bibliography

[1] Danish Energy Association. Dansk Elforsyning Statistik 2009, 2010.

[2] Peder Bacher, Anders Thavlov, and Henrik Madsen. Models for Energy Performance Analysis. Technical report, IMM, 2010.

[3] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st edition, 2007.

[4] John W. Brewer. Kronecker Products and Matrix Calculus in System Theory. *IEEE Transactions on Circuits and Systems*, 25(9):772–791, September 1978.

[5] Philip Anton de Saint-Aubain. Adaptive Load Forecasting. Master's thesis, 2011.

[6] Lars Elden, Linde Wittmeyer-Koch, and Hans Bruun Nielsen. *Introduction to Numerical Computation*. Studentlitteratur AB, 1st edition, 2004.

[7] Energinet.dk. Regulation A: Principles for the electricity market, 2007.

[8] Energinet.dk. Regulation C2: The balancing market and balance settlement, 2008.

[9] D. Hammerstrom et al. Pacific Northwest GridWise^{TM} Testbed Demonstration Projects, Part I. Olympic Peninsula Project. National Technical Information Service, U.S. Department of Commerce, October 2007. Online: `https://svn.pnl.gov/olypen`.

[10] Chong H. K. Goh and Jay Apt. Consumer Strategies for Controlling Electric Water Heaters under Dynamic Pricing. *Carnegie Mellon Electricity Industry Center Working Paper*, CEIC-04-02, 2004.

[11] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxfort University Press, third edition, 2011.

[12] Erick Herbin. *Polycopiés de cours de l'École Centrale Paris: Probabilitées*. 1st edition, 2007.

[13] Petros Ioannou and Jing Sun. *Robust Adaptive Control*. Prentice Hall, 1st edition, 1996.

[14] Henrik Madsen. *Modelling Non-Linear and Non-Stationary Time Series*. 1st edition, 2000.

[15] Henrik Madsen. *Time Series Analysis*. Chapman & Hall, 2008.

[16] K. Madsen and H. B. Nielsen. Supplementary Notes for Course 02611 Optimization and Data Fitting. `http://www.imm.dtu.dk/courses/02611/SN.pdf`, 2002.

[17] Casper Falkenstrøm Mieritz. Aggregate Modeling and Simulation of Price. Master's thesis, 2010.

[18] H. B. Nielsen. IMMOPTIBOX. `http://www.imm.dtu.dk/~hbn/immoptibox`, 2005.

[19] P. Nyeng and J. Ostergaard. Information and Communications Systems for Control-by-Price of Distributed Energy Resources and Flexible Demand. *Smart Grid, IEEE Transactions on*, March 2011.

[20] Preben Nyeng, Casper F. Mieritz, and Jacob Østergaard. Modeling and Simulation of Power System Balancing by Distributed Energy Resources and Flexible Demand. 2010.

[21] Fairey P. and Parker D. A Review Of Hot Water Draw Profiles Used In Performance Analysis Of Residential Domestic Hot Water Systems. Technical report, Florida Solar Energy Center/University of Central Florida, July 2004.

[22] Hanchuan Peng, Fuhui Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency,max-relevance, and min-redundancy. *Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, Aug. 2005.

[23] Joe A. Short, David G. Infield, and Leon L. Freris. Stabilization of Grid Frequency Through Dynamic Demand Control. *IEEE Transactions on Power systems*, 2007.

[24] SolarAnywhere. Global Sun Irradiance Data. `http://www.solaranywhere.com`, April 2011.

[25] Danmarks Statistik. Beboede Boliger KVM Og Antal Personer DK, 2010. `http://www.statistikbanken.dk`, Feb. 2011, 2010.

[26] Jan Bo Yang, Ka Quan Shen, Chong-Jin Ong, and Xiao-Ping Li. Feature Selection for MLP Neural Network: The Use of Random Permutation of Probabilistic Outputs. *IEEE Transactions on Neural Networks*, 20:1911–1922, 2009.