Sustainable Energy, Grids and Networks 10 (2017) 75–83

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan

Demand response evaluation and forecasting — Methods and results from the EcoGrid EU experiment[☆]

Emil Mahler Larsen [a,*], Pierre Pinson [b], Fabian Leimgruber [c], Florian Judex [c]

[a] *Danish Energy Association, Denmark*

[b] *Department of Electrical Engineering, Technical University of Denmark, Denmark*

[c] *Energy Department, Austrian Institute of Technology, Austria*

ARTICLE INFO

Article history:
Received 28 September 2016
Available online 18 March 2017

Keywords:
Demand response (DR)
Real-time pricing
Demand forecasting
Smart grid

ABSTRACT

Understanding electricity consumers participating in new demand response schemes is important for investment decisions and the design and operation of electricity markets. Important metrics include peak response, time to peak response, energy delivered, ramping, and how the response changes with respect to external conditions. Such characteristics dictate the services DR is capable of offering, like primary frequency reserves, peak load shaving, and system balancing. In this paper, we develop methods to characterise price-responsive demand from the EcoGrid EU demonstration in a way that was bid into a real-time market. EcoGrid EU is a smart grid experiment with 1900 residential customers who are equipped with smart meters and automated devices reacting to five-minute electricity pricing. Customers are grouped and analysed according to the manufacturer that controlled devices. A number of advanced statistical models are used to show significant flexibility in the load, peaking at 27% for the best performing groups.

1. Introduction

Interest in Demand Response (DR) has grown in recent years as system operators look for new tools to meet the needs of a rapidly changing power system. Changes include increased production from renewables, tighter market coupling, and a surge in decentralised production and consumption from photovoltaics (PV) and electric vehicles. The changing needs of the power system can broadly benefit from DR in two ways: through emergency use, where a reliable reduction in demand is needed during infrequent critical periods, and through economic use, where demand exhibits continuous flexibility to bring down average costs in the power system.

There are many ways of changing consumption patterns, but dynamic tariffs in particular are gaining interest due to their potential to respond quickly to fluctuating production from renewable energy sources (RES) [1]. Indirect control is one such dynamic tariff that uses an incentive signal, e.g. a real-time price, to influence the load. Indirect control does not require an exact response from any one customer, but with a large number of loads that exhibit somewhat similar behaviour, a statistically likely response can be forecast [2]. The value of indirect control therefore depends heavily on being able to accurately foresee its response to the incentive signal, which has previously been proven complicated [3].

The challenge of determining how much DR there is in a load has previously been done using baseline profiling [4–7]. Baseline profiling requires a prediction of the load under non-DR conditions which is then subtracted from the observed consumption under a DR event. A key drawback of existing methods is the need for data from non-DR days, data for which will not always exist, or may be unreliable since new equipment and interaction with customers can make non-DR data unrepresentative. Existing baseline methods are also unsuitable for evaluating fast moving DR that is conditional on a wide range of historical prices and price forecasts. Existing methods typically look a load curtailment, while we consider both increases and decreases in consumption due to decreases and increases in real-time pricing. Finally, existing methods may be susceptible to overfitting, and may rely on just one or two dozen observations per parameter.

Demand forecasting literature is a well developed area that is useful in predicting the price-elasticity of a load, e.g. see [8,9], but

☆ This work was partly supported by the European Commission through the project EcoGrid EU (grant ENER/FP7/268199).

* Corresponding author.
E-mail address: eml@danskenergi.dk (E.M. Larsen).

http://dx.doi.org/10.1016/j.segan.2017.03.001
2352-4677/

many modern approaches involve black box schemes like artificial neural networks (ANN), that obscure our understanding of the dynamics. Therefore, to get a full understanding of the controllable resources, we disaggregate the load into its constituent parts. Flexibility can then be interpreted in a useful way, so that it can be exploited for use in different services, or bid into a market. There are no existing methods that are appropriate for evaluating and integrating residential DR into a balancing market, in particular when data for non-DR days is unavailable, which was the task we had to achieve in the large scale smart grid experiment, EcoGrid EU.

We cultivate modelling approaches that determine how much DR a load is capable of delivering in terms of power and energy, and under what external conditions, e.g. ambient temperature. More specifically, DR is characterised in terms of peak response, time to peak response, energy delivered and ramping. Our primary motivation was to compare the performance of groups of houses with different hardware and software that receive real-time pricing. The tools developed were used to give feedback to hardware and software manufacturers so that their algorithms could be improved. Our second objective was to apply these attributes in the constraints of a balancing market so that the load could be controlled. With a balancing market scheduling DR, we sought to validate our method by comparing the observed response to the scheduled DR. Aside from the approach, an additional contribution are the state-of-the-art estimates of residential flexibility used in a five minute balancing market.

The paper is structured as follows: Section 2 investigates existing approaches for evaluating the success of DR, as well as previous work that was relevant when developing our own methodology. Section 3 introduces the experimental setup and the data gathered from the demonstration. Section 4 describes the models developed to analyse the DR activated during the demonstration. Section 5 presents DR analysis for different groups and results from real-time forecasting in the demonstration. Section 6 discusses uncertainty, future work and concludes.

2. Demand modelling

There are several lines or research that are relevant when assessing a DR program, including previous experimental studies, load forecasting research, and energy disaggregation research. Previous experimental analyses have looked at different types of DR, like critical-peak pricing (CPP) and time-of-use (TOU) pricing, and often consider human demographics and behaviour as an impact on DR. Forecasting literature has a widespread use in operation of power systems and offers a deeper insight into the statistical tools available, with a greater focus on weather conditions, calendar effects and economic variables. Energy disaggregation research is an up and coming area driven by new sources of data, like high resolution smart meter data from thousands or millions of customers.

2.1. Previous DR studies

The study of residential loads responding to prices goes all the way back to the 1970s and many fundamental aspects, like accounting for the time of day and ambient temperature in a statistical model, remain in use today. Studies have also included home type, size, income and smart thermostat ownership as model inputs [10,11], but residential DR studies have not been able to give concrete numbers in terms of power and energy the load is capable of delivering.

For medium and large commercial and industrial loads, baseline methods are widely used to determine financial settlement for participating customers. The baseline is simply the prediction of consumption under the assumption that no DR was present. The baseline is then subtracted from the observed consumption to determine the amount of load shed the customer was able to deliver during a critical period. Baseline models are created by regressing on historical data before DR events. This has been done with hourly interval data [4] and 15 min interval data [6]. In the latter case, it was observed that including parameters for load shed directly into the model did not give a reliable result, possibly because the model was too primitive or due to over parametrisation. Another approach is to average consumption for just 5–10 days before a DR event [5], yet such few observations may mean that this approach lacks robustness.

2.2. Forecasting approaches

Short-term load forecasting presents a number of useful tools that can predict how load changes with respect to price. Classical approaches to solving hour-ahead and day-ahead load forecasting problems include time-series methods like auto-regressive integrated moving average (ARIMA) models and exponential smoothing, also including geographical factors [12] and seasonal variations [13].

Recent advances in forecasting methods include spline-based methods [9], which avoid over-parametrisation by relying on a handful of splines to describe the baseload, although authors in [9] noted that some fidelity was lost during peak load periods. This work was applied to a price responsive load in New York, with parameters for price and, in theory, these parameters should allow a full evaluation of the DR volume, although this was not explored in practice.

Other modern advances in forecasting include multivariate state-space models [14], which feature time-varying regression coefficients that may be useful for analysing DR. Semi-parametric methods to predict the contribution of load from some non-linear variables [15] may also be useful for DR volume evaluation, although [15] did not apply the methodology to a price-responsive load.

Machine learning approaches like artificial neural networks (ANN) are also popular for forecasting, with positive results reported in [16]. The benefit from ANN includes being able to capture unspecified non-linear relationships between external variables like weather. It is likely that such an approach becomes increasingly valuable as demand becomes more non-linear and volatile with new external incentives like price and the growth of distributed energy resources (DERs). ANNs have, however, been criticised for leading to over-parametrised models [17] and do not necessarily outperform linear regression models [18]. From a DR evaluation perspective, ANN's black box form makes picking out price influences complex, especially when bidding a price response into an electricity market.

2.3. Energy disaggregation

Energy disaggregation has gained interest as automatic metering infrastructure becomes ubiquitous in many countries. Energy disaggregation tools can be used to see beyond the meter and uncover which devices are turned on despite only seeing a noisy, aggregated view of the load. The stated goal of disaggregation is to better understand the load and make well-targeted energy efficiency plans.

Of particular relevance to our study is the success in [19] of detecting air-conditioning use from 1-min interval smart meter data. Methods that rely on a dictionary of devices describing the real and reactive power each consumes have also previously been developed [20]. Grey-box, Markovian stochastic, Bayesian and logistic adoption models are also promising ways of identifying human behaviour and price-responsive devices in metering data [21]. However, disaggregation techniques are not conclusively proven with external influences or with variable speed devices.

3. Data and experimental setup

Data for this work comes from EcoGrid EU, which is an indirect control experiment on the Danish island of Bornholm. The experiment uses a market framework to schedule regulating power from conventional generation and DR to meet the real-time balancing needs of the system, needs which are growing due to increased wind power production. DR is activated with a five minute price signal sent to 1900 houses. Houses are equipped with smart meters that collect data every 5 min, which in turn is used as an input to forecast demand in real-time. A mathematical introduction to price generation in the EcoGrid EU market can be found in [22].

Demonstration houses are fitted with a wide range of small-scale DERs like heat pumps (HPs), resistive electric heating (EH) and hot water boilers. Heating is the main source of flexibility and is expected to grow due to electrification of heating systems. Flexibility of such installations comes from houses retaining heat and heating devices therefore only needing to be turned on sporadically to meet customer comfort requirements. Some households also had different sources of heating; in the heat pump and resistive electric heating groups, 22% and 28% respectively had an alternative source of heat, predominantly wood stoves. There was no reliable way of determining when these sources were in use and, as such, they are treated as an additional source of load correlated with ambient temperature. 12% of households were fitted with PVs, but its production is included in the aggregate metering data (i.e. net metering) and was therefore treated like any other source of uncontrolled load.

Households were enrolled through a outreach program by the local distribution grid operator and electricity retailer. Activities included setting up stalls at fairs and other local events, and contact through telephone, mail and local TV. Households were not paid to participate, but were offered the chance to be a part of an exciting project that supports renewable energy, the chance to get new equipment to control their devices and the possibility of a lower electricity bill. Ultimately, half of all households with a heat pump on Bornholm were recruited.

Table 1 shows the group composition that is the basis of our analysis. The IBM HP, IBM EH and TNO HP groups all use the same smart thermostat hardware, while the Siemens EH group uses different smart thermostat hardware. Different control algorithms were used in each group.

The reference group was designed be a control group for all other groups. It included households with heat pumps, resistive electric heating and manual customers with other sources of heating—roughly a third of each customer type. However, it was not a representative control group for any test group. An example of the difference between the reference and Siemens EH groups is shown in Fig. 1. Here, the prices sent to customers is shown on the top plot, and the consumption of the reference group and Siemens EH group are on the bottom plot. The Fourier time-series on the bottom plot represents behaviour due to the group demographics, which is not the same. The methodology section describes how to extract this Fourier series. Several customer surveys showed that HP owners have significantly higher incomes, much larger houses, are more likely to own their homes, and are younger than customers who use EH, which explains some of the demographic differences between groups. Unrepresentative reference groups are a key driver for the development of a comprehensive modelling approach.

Smart meter data was filtered to remove houses with fewer than 80% of measurements, leaving 1707 out of 1900 houses. It is never the case that all smart meters report consumption in each time period, so consumption was normalised to the full population of each group. Periods where fewer than 85% of houses reported

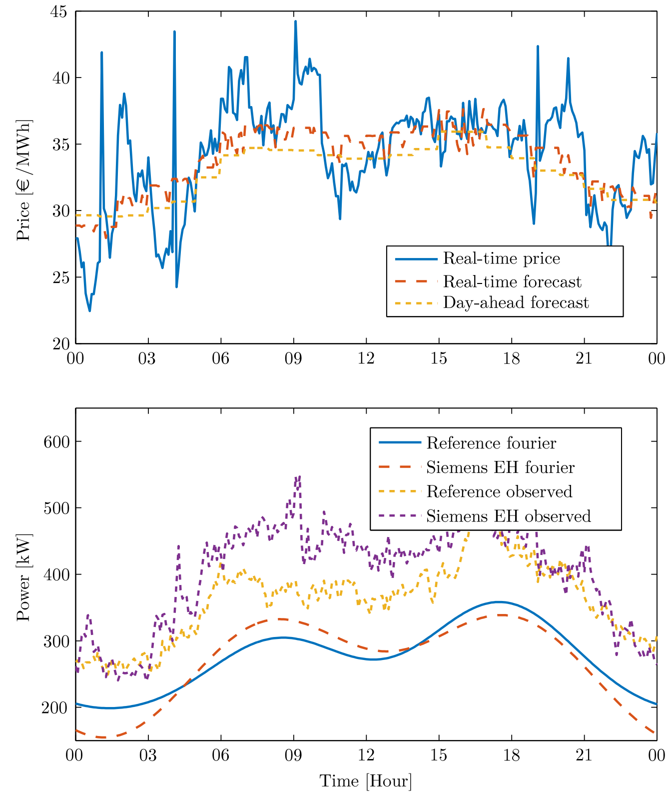

Fig. 1. Single day example of the reference group and a group of automated houses showing profiles that are too different for direct comparison.

their consumption were left out, leaving 40,000 observations for each group. The analysis period covers September 22nd 2014 to February 8th 2015, with ambient temperatures spanning −10.6 °C to 18.0 °C.

4. Methodology

To evaluate the DR potential of each group, we started with a standard linear model with additive terms to describe the entire (aggregate) load, including flexible and non-flexible demand. Linear models have been previously noted to be competitive with newer methods, like probabilistic linear regression and ANN [18]. Unlike some newer methods, parameters of a linear model offer a clear and simple description of how different variables contribute to the load.

A linear model is essentially an extrapolation that approximates reality but does not capture the exact events that occurred. This is especially true for individual DER scheduling algorithms that may respond unexpectedly when given a combination of prices that has never occurred before. However, our work, builds on the hypothesis that a large enough population of DERs will deliver a response that is proportional in some way to the input signal and a linear model is capable of capturing the average response given a wide range of random price signals and response.

We expanded the linear model with non-linear terms for prices, which bound the response, since linear parameters otherwise predict an infinite response to prices, given infinite prices.

A key challenge with linear models is collinearity, which means that different external variables, for example solar irradiance and ambient temperature, are correlated and explain the same outcome. This is particularly troublesome for the day-ahead prices, which came directly from the Nord Pool spot market and were not controlled in the experimental setup. Day-ahead pricing is correlated with the inflexible part of the load, while the flexible

Table 1
Summary of EcoGrid EU customer groups.

Reference	253 households with smart meters	No pricing information	279/621 kW average/peak load
Manual	455 households with smart meters	Receive real-time pricing, but must alter consumption manually	287/750 kW average/peak load
IBM HP	195 households with smart meters and automation equipment	Heat pumps react autonomously to prices	317/710 kW average/peak load
IBM EH	322 households with smart meters and automation equipment	Resistive electric heating reacts autonomously to prices	394/854 kW average/peak load
TNO HP	84 households with smart meters and automation equipment	Heat pumps react autonomously to prices	123/293 kW average/peak load
PV Siemens EH	398 households with smart meters and automation equipment	Resistive electric heating, water boilers and controllable PV react to aggregator control	300/810 kW average/peak load

part of the load reacts to the subsequent pricing. Separating active DR from a correlation that would be there anyway, i.e. cause and effect, is therefore difficult. Yet it is necessary to determine the price sensitivity when bidding into a day-ahead market, i.e. [23]. Likewise, determining the price sensitivity to real-time prices is also needed to bid into the balancing market, i.e. [22].

A second model that relies upon differenced variables was applied and found to be less susceptible to collinearity. Our experience taught us that the differenced model had a lower short-term forecast error, which made it well suited to trying to control the load every five minutes in the experiment. It also appears to capture the fast moving price-dynamics more accurately.

Despite the benefits of the second model, the standard model is still needed because it gives a different result. The standard model gives the absolute value for DR and is more accurate for long-term forecasting. The differenced model tells us the maximum change in DR in any five minute interval, which is needed when scheduling DR in an operational environment, like the EcoGrid EU experiment.

4.1. Model 1: a linear model

Initially, demand is conceptionally split into a component dependent on external variables and a component that is dependent on external variables and prices using the notation from [24], i.e.

$$c_t = f\left(\tilde{z}_t\right) + g\left(\tilde{\lambda}_t, \tilde{z}_t\right) \tag{1}$$

with

$$\tilde{\lambda}_t = \begin{bmatrix} \lambda^D_{t+u^D_\lambda}, \ldots, \lambda^D_{t-n^D_\lambda} \\ \lambda^R_{t+u^R_\lambda}, \ldots, \lambda^R_{t-n^R_\lambda} \end{bmatrix}$$

$$\tilde{z}_t = \left[z_{t+u_z}, \ldots, z_{t-n_z}\right]$$

where n^D_λ, n^R_λ and n_z are a finite number of lagged values of day-ahead price, λ^D, real-time price, λ^R, and external variables, z. There are also u^D_λ, u^R_λ and u_z forecast values, which capture the scheduling dynamics of DERs. All variables are first centred by subtracting their mean value, while price forecasts are converted to relative prices each time a new price forecast is made. The day-ahead price $\lambda^{\mathrm{D,raw}}$ is normalised to a value between 0 and 1, so that

$$\lambda^{\mathrm{D}}_t = \frac{\lambda^{\mathrm{D,raw}}_t - \tilde{\lambda}^{\mathrm{D,min}}_t}{\tilde{\lambda}^{\mathrm{D,max}}_t - \tilde{\lambda}^{\mathrm{D,min}}_t}. \tag{2}$$

Real-time prices and real-time forecasts are also converted to relative prices, but with respect to the absolute day-ahead price, i.e.

$$\lambda^{\mathrm{R}}_t = \lambda^{\mathrm{R,raw}}_t - \lambda^{\mathrm{D,raw}}_t \tag{3}$$

where the absolute day-ahead price, $\lambda^{\mathrm{D,raw}}_t$, is the price that was forecast for time t the day before (so priced at $t - 24$ h).

External variables, z, include weather data such as wind speed, solar irradiance, temperature, Φ, and a base load term, y. Exponentially weighted smoothing is performed on weather variables, i.e.

$$z_t = \frac{\sum_{i=1}^{n_z} \alpha^{i-1} z^{\mathrm{raw}}_{t-i+1}}{\sum_{i=1}^{n_z} \alpha^{i-1}}. \tag{4}$$

The steps to find the number of lags, i, and the smoothing factor, α, are found experimentally (as described in [18]), where the lowest error is the determining factor.

A non-linear transformation is also applied to temperature, whereby it is modelled as a third-order polynomial, as has previously been proven successful in [25], such that

$$\tilde{\Phi}_{t-n} = \left[\Phi_{t-n} + \Phi^2_{t-n} + \Phi^3_{t-n}\right] \tag{5}$$

where $\{n \in \mathbb{Z} \mid 0 < n \leqslant n_z\}$. Modelling temperature as a third order polynomial introduces more coefficients than a model without this non-linear transformation.

The base load, y, is a Fourier series that describes demand due to the time of day, day of the week, and day of the month [26], such that, for a given time t,

$$y_t = \sum_{k=1}^{K} a_k \sin\left(\frac{2\pi kt}{T}\right) + b_k \cos\left(\frac{2\pi kt}{T}\right) \tag{6}$$

where T must be suitably large to cover different seasonal variations (for example 288 when capturing trends of different hours of the day using five minute data) and K is increased until enough high-resolution detail is captured. The Fourier terms includes different subsets, including a series that is consistent for the days of the week (Monday to Sunday), equivalent to $T = 2016$, as well as a monthly Fourier ($T = 8928$). K is adjusted so that the fastest components (both sine and cosine) have a frequency of approximately half a day. Applying a monthly Fourier series to a continuous time-series with 28, 30 and 31-day months causes a loss of synchronicity, yet the monthly Fourier series did appear to reduce model error. In any case, the weekly Fourier series captures a significant level of detail.

Ideally, several years of data would be used, with a Fourier series for each month of the year. However, this would still cause problems when encountering leap years. Alternatively, a model with terms for each hour of the day and terms for each month of the year, might be appropriate, given several years of data, as described [18].

Additional terms are added to capture the interaction between the base load and temperature, $y\Phi$, and the price and temperature, $\lambda\Phi$, included in the array of variables $\tilde{\chi}$. The full model for demand can be expressed in linear model form, i.e.

$$c_t = \tilde{\lambda}_t^\top \tilde{\theta}_\lambda + \tilde{z}_t^\top \tilde{\theta}_z + \tilde{\chi}_t^\top \tilde{\theta}_\chi + \epsilon_t = \boldsymbol{x}_t^\top \boldsymbol{\theta} + \epsilon_t \tag{7}$$

where ϵ is Gaussian noise with zero mean and finite variance. The resulting parameters for price are time-invariant, although its possible that additional interaction terms between the time and price could solve this issue. We did not implement this because DR was expected to come from automated DERs whose performance only depended on ambient temperature.

Relying on a conventional least squares regression with five minute data and several lags for external variables can lead to a model with hundreds of parameters and suspicions of overfitting, even in light of the 40,000 observations that the model is fitted to. As a result, we found the only way to solve the model and get reasonable parameter estimates was to minimise the residual sum of squares while shrinking parameters using the Lasso penalisation [27]. The objective of this is

$$\min \sum_{t=1}^{T} \left(c_t - \boldsymbol{x}_t^\top \boldsymbol{\theta}\right)^2 + \eta\, |\boldsymbol{\theta}| \tag{8}$$

where η is the tuning parameter and is found using a 10-fold cross-validation (CV) routine, minimising MSE_{CV} (MSE). Given a model with 200 input parameters, the Lasso penalisation gave just 50–70 non-zero parameters for each group using both the standard and the differenced model.

4.2. Model 2: a differenced model

The differenced model relates how consumption changes with respect to how the variables change through time, i.e. gradients of external variables like price and temperature affect the gradient of consumption. Conceptionally, the split becomes

$$\dot{c}_t = f\left(\dot{\tilde{z}}_t\right) + g\left(\dot{\tilde{\lambda}}_t, \dot{\tilde{z}}_t\right) \tag{9}$$

where $\dot{c}_t$, $\dot{z}_t$ and $\dot{\lambda}_t$ are the change in consumption, external variables and prices at time t respectively. The interaction terms between temperature and price should be interpreted as how fast the change in consumption occurs due to how fast the price and temperature are changing. As a result, to better understand ambient temperature effects, the parameters of the linear model are found in Eq. (8) for three different ambient temperature bins of equal population, which means that the parameters are found with 13,000 observations each time. From a classical time-series analysis perspective, differencing ensures stationarity, as it removes many collinearity problems. The downside of differencing the model is that it removes useful information in our case, since the parameters no longer relate to the absolute effects present.

Table 2
10-fold cross-validation by group.

	MSE (kW)	MSE_{CV} (kW)
Reference	112	111
Manual	122	120
IBM HP	226	229
IBM EH	205	204
TNO HP	264	266
Siemens EH	66	66

4.3. Cross validation

To ensure that our models are not overfitted, and to test the data for consistency, a cross-validation approach was used. Specifically, a ten-fold cross-validation was employed. The data was randomly divided into 10 approximately equally sized subsets or folds. For each fold (validation fold), the remaining other 9 folds were used to first train the model. The model was then applied to the observations in the validation fold, resulting in a mean squared error for the validation fold. This procedure is repeated for the other 9 folds respectively, resulting in 10 estimates of the test error in total. These estimates are averaged, resulting in the CV estimate MSE_{CV}. To assess model performance with respect to time, this cross-validation estimate is compared to the MSE obtained by fitting the model to the full data. The relative difference of the full-data MSE and the MSE estimated by CV is then used as a measure of model consistency across different random portions of the data. Table 2 shows the CV outcome for the standard model. The almost identical results for 10-fold MSE and full-sample MSE means that the model is consistent over time and less likely to be overfitted.

4.4. Non-linear terms for price

In an operational environment, specifically the market within which DR in EcoGrid EU is controlled, bounds for the price-response must be given for each five minute period. Without bounds, it is impossible to bid a sensible level of DR (and its maximum ramp-rate) into the power market. The relationship between price and demand has previously been observed to be non-linear [24] and this can be modelled by redefining the price terms in a generalised logistic function that is centred around zero. Non-linear terms for price were only applied to the relative real-time price, which is the quantity bid into the real-time market (and where practical bounds are needed). To find non-linear price terms, the residuals from both Models I and II were combined with the price response according to the Eq. (7), i.e.

$$\epsilon_t^* = \epsilon_t + \lambda_t \theta_\lambda. \tag{10}$$

The residuals ϵ^* should now include the predicted linear DR and non-linear components too. The objective to find the logistic function parameters is given by

$$\min \sum_{t=1}^{T} \left(\epsilon_t^* - \left(\sum_{n=1}^{N} -\frac{A_n}{2} + \frac{A_n}{1 + e^{-\varepsilon_n \tilde{\lambda}_n}}\right)\right)^2 \tag{11}$$

where the N most important lags are chosen, which are those with the greatest parameter estimate (i.e. greatest price elasticity). This was typically the first three lags. A range of commercial solvers were experimented with, showing that models with a maximum of three logistic functions could be solved for.

In Eq. (11), $\tilde{\lambda}$ is an array that includes n lags and t time-steps. $A_n/2$ is the maximum amplitude of the response due to price lag n, while ε_n describes the price-elasticity. Parameters are found by minimising the sum of square errors using the Levenberg–Marquardt algorithm [28]. The starting estimates for ε_n and A_n were found by first considering the derivative of the logistic

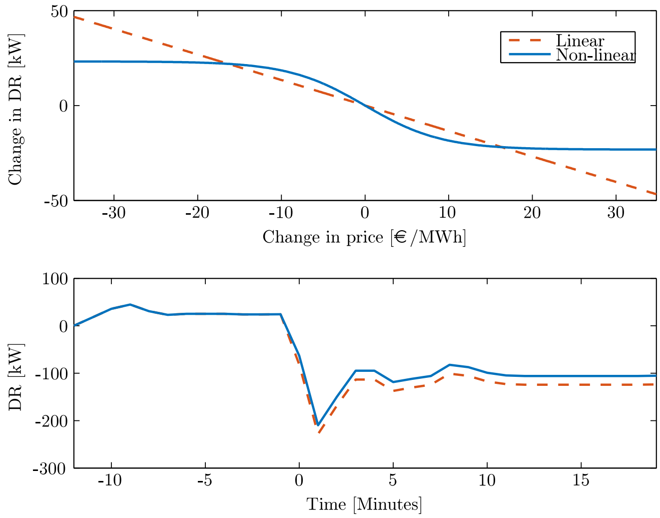

Fig. 2. A comparison of a linear parameter for price and a novel non-linear model for price. The response of the non-linear model (bottom plot) is bounded for very high electricity prices. The example in this case applies to the differenced model.

function with respect to price. Then, as the price tends to zero, the derivative tends to $\frac{A_n \varepsilon_n}{4}$. Assuming that ε_n is equivalent to the linear parameter for price (for small prices) means that A_n should be 4. Therefore, the linear parameters for price and 4 were used for starting estimates of ε_n and A_n respectively. A sensitivity analysis with several starting estimates would be desirable for future work, to investigate the possibilities of ending in a local minimum, as is possible with non-linear problems.

Fig. 2 illustrates the impact of non-linear terms for price, where the top plot shows the boundary of DR, and the bottom plot shows the accumulated impact on the finite impulse response (FIR), that is the change in load due to price, through time, when given the data obtained in the EcoGrid EU demonstration. If non-bounded price terms are used, then the linear prediction (dashed red line in top plot) will tend to infinity if the prices tend to infinity. The non-linear model enforces bounds on this response, beyond which consumption cannot change further in the presence of extreme prices.

Non-linear terms reduce DR for each group by 2%–16% compared with linear terms for DR, given the maximum prices seen, while the mean absolute percentage error (MAPE) is only fractionally reduced (0.05%). In a differenced model, the initial change to price gives the biggest DR load contribution, since it impacts all subsequent price lags. In the standard model, the biggest change to price is assumed to be largest parameter for price, which happens 10–30 min after a price change.

4.5. Deriving useful DR measures

The statistical models developed can now be used to bid DR into day-ahead and real-time markets assuming standard quadratic programming structures, as is in use today [29]. For the EcoGrid EU demonstration, this meant bidding the price elasticity of the entire population into a real-time balancing market, which are the parameters associated with λ^R.

To evaluate DR, more specifically to determine peak response, energy delivered and ramping, the statistical models with non-linear terms for price where filled with observations and meteorological data and calculated as follows. For the peak real-time DR delivered

$$\text{Maximum DR} = \max \frac{|A_n|}{2}. \tag{12}$$

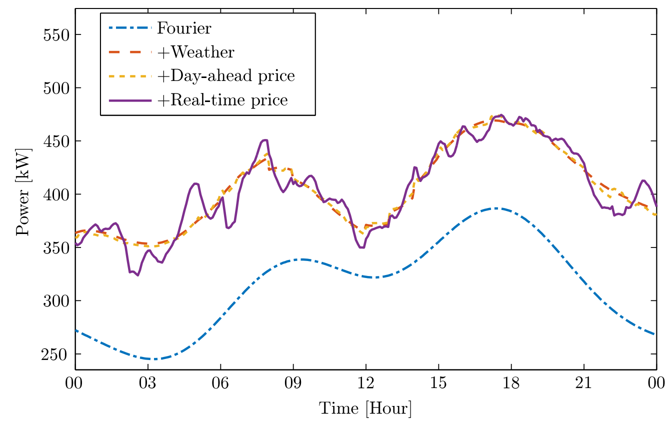

Fig. 3. Load contributions according to the linear model. Each contribution to the load is added to the previous one.

For the initial ramping, we consider only the price lag for the instantaneous price, i.e.,

$$\text{Initial ramping} = \frac{|A_1|}{2}. \tag{13}$$

The maximum DR and initial ramping do not depend on the price elasticity, i.e. the maximum amplitude of the logistic function is independent of its gradient.

The first volume result in Table 5 is defined as the sum of the absolute price impulse over the entire period analysed. When there is no price variation, then the volume is zero. For the maximum energy delivered in one hour (volume max in Table 5), we integrate over the first 12 values of the FIR, while the time to peak response (Table 6) is simply the time, n, for which A_n is largest.

The day-ahead DR component of the model does not employ logistic terms and is calculated as the maximum of the time-series of day-ahead price multiplied by the parameters for day-ahead price.

5. Results

The power and energy that can be delivered by each group is presented here, using the two modelling approaches (non-differenced and differenced) developed. We also present the outcome of forecasting the load using the differenced model. Fig. 3 shows the outcome of the non-difference model, with Fourier terms, weather, day-ahead and real-time price effects cumulatively added "on top" of each other.

5.1. Group FIR visualisation

The FIRs related to price parameters from the differenced model are shown for the different automated groups in Fig. 4 for three different temperature bins. In this figure, at time zero, the relative real-time price changes from 0 to the maximum price change observed in the experiment. The load responds immediately, reaching a peak, then rebounding, then arriving at a new steady state after approximately 50 min. Before the price change, real-time price forecasts cause consumption to increase, causing load to shift. This figure could also be given as a mirror image, showing load increase for a reduction in price. The Siemens EH group exhibited the largest and fastest response to prices, while the IBM groups exhibited a slower response to the price increase. We observed no change in consumption for TNO HP houses. For the warmest periods, which averaged 11.3 °C, the Siemens EH group exhibited significant DR, while IBM EH houses also exhibited a small but noticeable change in consumption. There was no DR from the other groups during warm conditions.

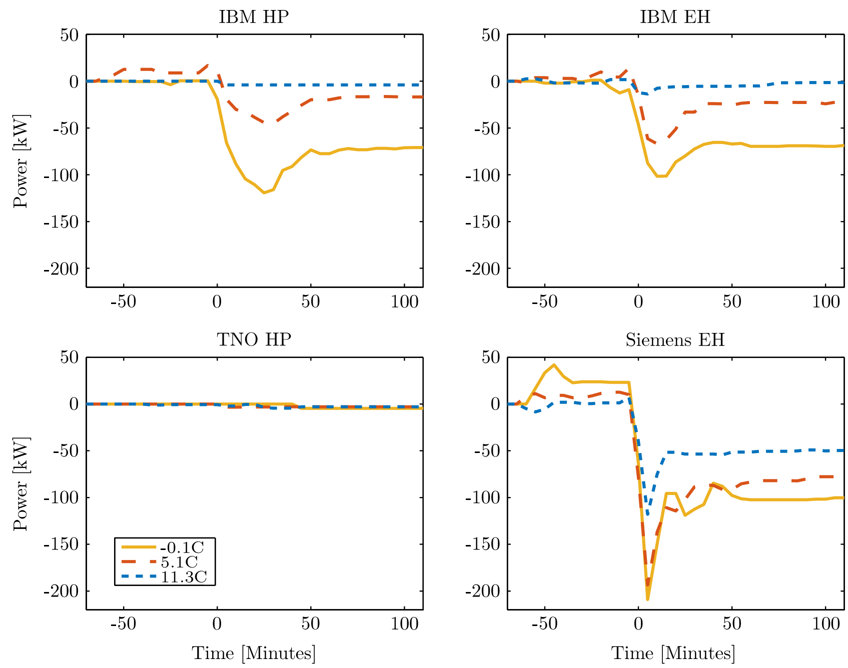

Fig. 4. The response to price for different groups, given the largest price change observed.

Table 3
Maximum real-time DR.

Group	DR (kW)	Normalised DR
Reference	0	0
Manual	4.0	0.005
IBM HP	190.1	0.268
IBM EH	114.4	0.134
TNO HP	0	0
Siemens EH	198.0	0.244

Table 4
Maximum day-ahead DR.

Group	DR (kW)	DR (differenced model) (kW)
Reference	−29.1	−25.6
Manual	−28.9	−24.0
IBM HP	7.3	10.6
IBM EH	−39.9	−30.4
TNO HP	−8.3	−4.5
Siemens EH	0	35.7

5.2. DR evaluation

Table 3 shows the peak DR according to the standard linear model, also normalised with respect to the peak consumption of each group. In absolute terms, the Siemens EH group exhibits the biggest DR, while when normalised for the peak consumption, the IBM HP group appears to give the biggest response, totalling 27% of peak load. The reference group shows no DR, as expected, while the manual group shows a very small amount of DR. No DR was detected for TNO HP houses.

In Table 4, the consumption due to day-ahead pricing is shown for the standard and differenced models. A negative sign indicates that consumption and price are positively correlated. For example, if the price goes down and the consumption is stated as −10 kW, then consumption decreased by 10 kW even though the price decreased. A higher value (i.e. non-negative) either means that the scheduling controller is more responsive to day-ahead price, or that the group has demand that occurs naturally at different times of the day than the majority of the demand that drives the day-ahead price. The differenced model shows less positive correlation across all groups, indicating that there is less collinearity between base load and price, i.e. in the differenced model, the changes in the day-ahead price are less correlated with the changes in the load. It is possible that other terms better describe changes in the load for the difference model, namely the Fourier terms. The benefit of the differenced model is especially noticeable for Siemens EH houses, which appears to exhibit the largest response to day-ahead pricing, followed by IBM HP houses.

Table 5 shows the load shifting characteristics of each group, based upon the price-related parameters in the differenced model.

Table 5
Real-time DR volume.

Group	Volume (kWh/h)	Volume normalised	Volume max (kWh)
Reference	0.4	0.010	−9.8
Manual	0.3	0.006	−1.2
IBM HP	4.9	0.115	93.0
IBM EH	3.9	0.078	73.5
TNO HP	0.1	0.008	1.2
Siemens EH	9.5	0.206	110.3

The first column shows the average DR volume shifted per hour throughout the test period. The second column normalises this value by the average energy consumed in each group. The final column states the maximum load shift in a single hour, when given the maximum price change observed. This is the area over the FIR curves shown in Fig. 4. In Table 5, a negative sign can indicate either an economically-irrational response or be due to the uncertainty in the model. The reference group, which does not receive prices, exhibits this economically-irrational, positive correlation. But it is one order of magnitude smaller than the test groups, suggesting that the uncertainty may also be one order of magnitude smaller than the DR activated. The average volume for Siemens EH houses appears to be twice that of IBM HP houses, but it is only 19% higher in terms of maximum volume. This can be attributed to the faster ramping abilities of the Siemens EH group, as shown in Table 6, which means its DR is exploited more by the market, since quick, small changes are activated more in the market tested.

Table 6
Ramping and time to peak response.

Group	Initial ramping (kW)	Time to peak response
Reference	−4.4	N/A
Manual	0	N/A
IBM HP	19.1	30 min
IBM EH	46.2	15 min
TNO HP	0	N/A
Siemens EH	63.4	10 min

5.3. Forecasting application

The differenced model with additional auto-regressive components was used to forecast aggregate consumption of all houses in real-time in the EcoGrid EU demonstration, using real-time meteorological forecasts from the Danish Meteorological Institute (DMI).

Forecasting was operational in the period March 10th 2015–April 30th 2015, with parameters fitted using data from September 22nd 2014 to February 8th 2015. The auto-regressive components were consumption lags, so that the full model implemented was

$$c_t = f\left(\tilde{z}_t\right) + g\left(\tilde{\lambda}_t, \tilde{z}_t\right) + h\left(\tilde{c}_{t-1}\right). \tag{14}$$

It should be noted that the consumption lags now include information about the currently activated DR. This model can therefore not be used to assess the entire volume potential.

The MAPE for one step ahead was 1.6%, compared with a persistence (where the load will be the same as it was five minutes ago) MAPE of 2%. Hindcasting gave MAPEs of 1.2% and 1.4% for models with and without price terms respectively. Hindcasting with the standard linear model without auto-regressive terms gave a MAPE of 5.5%. Fig. 5 shows how the MAPE evolves with forecast horizon. We witnessed large errors for longer forecast horizons due to smart meter aggregation variability. For one five minute period, 250 smart meters might report in, with their load normalised to 1900 houses (the full population). In the next five minute period, 1700 houses might report in, which also causes the earlier observations to change. Smart meters contain a Sim card and report consumption over standard mobile phone networks. Varying delays are caused by cost and bandwidth limitations and patchy signals for some houses. The average jump in meter data for any five minute time period is 7% and, in the worst periods, this reaches 25% of the load; the results given above are for where the recent historical meter readings for the normalised aggregate population did not change by more than 0.5%. With such volatility, the auto-regressive contribution to the load forecast was extremely misleading, causing large errors for longer forecast horizons. In a full-scale roll-out, grid measurements such as system frequency could be used instead of smart meter data in our models. If smart meters were still to be relied upon, then the most representative smart meters should be identified, made robust and be relied upon.

In spite of poor long-horizon forecasting performance due to smart meter collection variability, the five minute ahead forecast was good enough to see a real-time response. Fig. 6 shows the change in DR requested by the market, binned into 20 kW groups, which is subsequently converted to a price based on the demand models created here and sent to customers. The smallest bin had only 10 observations, while the largest bins had hundreds of observations. The change in load observed shows the raw consumption data that includes all other sources of uncertainty, like baseload, weather and human behaviour. Despite these sources of noise, there appears to be a clear trend where demand is following the price. For the largest requests in a reduction in consumption, a reduction in consumption occurred 80% of the time, while for the largest requests for an increase in consumption, an increase in consumption occurred 100% of the time.

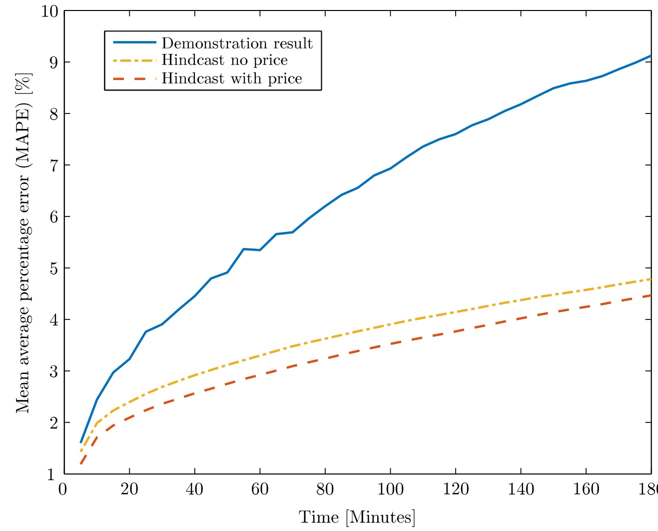

Fig. 5. Mean absolute percentage error (MAPE) from the experiment and from hindcasting on historical data.

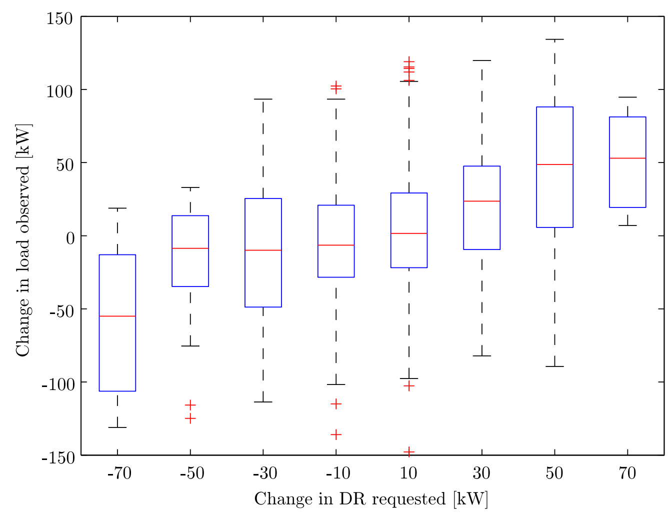

Fig. 6. Change in load observed, binned for different levels of DR activation based upon the forecasting model. Each boxplot shows the median as a red bar and the blue box showing the two middle quantiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.4. Result uncertainty

Uncertainty is clearly present in our results and best observed in Table 5, where the reference group appears to exhibit small amounts of DR, despite not receiving the real-time price. This is due to spurious correlation, but it is an order of magnitude smaller than the automated groups.

There are many sources of error in a residential DR system, the most significant of which is likely structural, i.e. models do not capture the full dynamics and all interactions in the load. Experimental errors include that the model is unaware when smart control equipment is off for maintenance or due to communication errors, which raises the possibility that the maximum DR potential is underestimated here. Weather observation and forecast errors are also present, since weather data comes from a single point that will not be representative of the entire population.

Previous DR error analyses have relied on standard errors [4]. Methods to derive standard errors from a Lasso regression have also been developed [30], yet the scientific community and leading journals are rejecting such significance tests since they are often interpreted incorrectly. In [3], standard errors for DR evaluation were identified as underestimating the uncertainty for a number of reasons. In our case, these include correlation between Fourier terms, temperature, and day-ahead pricing. Regression residuals are also autocorrelated, especially since we do not include auto-regressive or moving-average terms in our models (these terms are only meaningful in a forecasting environment). Regression residuals for baseline models have also been identified as being heteroscedastic; i.e. the error is dependent on external variables like the time. Leave one out cross validation (LOOCV) is a good solution for estimating the true level of uncertainty in a baseline framework with observations averaged over several hours [6]. It is, however, far too computationally intensive when dealing with the five minute interval data and the 40,000 observations that we rely upon. We are therefore left without an appropriate toolbox to assess uncertainty, and this should be a major focus of future work in this field.

6. Conclusion

We have determined that our load is capable of delivering significant DR, where up to 27% of the peak load is flexible for the best performing groups. The software algorithm for the TNO HP group, which did not exhibit any DR, was subsequently changed following our findings and a response to price was observed in a later time period. This highlights how such statistical models could be used in the future; Revenue from following dynamic pricing may not be enough to justify the high investment cost needed to fit houses with automation equipment, which raises the possibility of capacity payments being made to support initial DR installations. The statistical models developed here could therefore be used to determine financial settlement based on peak response and average energy provided. We did not see noticeable DR from the manual customers, but this could be due to the design of the experiment: customers were guaranteed an electricity bill no higher than if they were not in the demonstration. A large scale roll-out of dynamic pricing would have no such guarantee and would penalise as well as incentivise customers, which would likely result in increased DR from customers with no automation equipment.

We spent some time investigating a differenced model that included non-derivative variables, but these variables would fall out of the model once the Lasso procedure was applied—i.e. the non-derivative variables were not correlated with the differenced load. A model that can combine aspects of both models would be desirable. Aside from such a model, future work should also look to advanced machine learning models, since our models likely ignore significant non-linearities in the load. This is especially relevant for analysing the DR potential of IBM HP, IBM EH and TNO HP groups, since the hardware used in these groups can only turn off heating devices, not on. In a large enough population, with a diverse range of internal states, this may not mean a significant asymmetrical response for thermostatic loads, but it warrants investigation nonetheless.

Acknowledgements

The authors thank our EcoGrid EU partners for supporting this work. The authors also thank DMI and ECMWF for weather forecast data.

References

[1] N. O'Connell, P. Pinson, H. Madsen, M. O'Malley, Benefits and challenges of electric demand response: A critical review, Renewable Sustainable Energy Rev. 39 (2014) 686–699.
[2] K. Heussen, S. You, B. Biegel, L.H. Hansen, K.B. Andersen, Indirect control for demand side management - A conceptual introduction, in: IEEE PES Innovative Smart Grid Technologies Conference Europe, 2012, pp. 1–8. http://dx.doi.org/10.1109/ISGTEurope.2012.6465858.
[3] J.L. Mathieu, D.S. Callaway, S. Kiliccote, Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing, in: IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 4332–4339. http://dx.doi.org/10.1109/CDC.2011.6160628.
[4] S.D. Braithwait, D.G. Hansen, J.D. Reaser, Load impact evaluation of California statewide demand bidding programs (DBP) for non-residential customers: Ex post and ex ante report, Tech. Rep., Christensen Associates Energy Consulting, Madison, WI, 2011.
[5] M.L. Goldberg, G.K. Agnew, Commission protocol development for demand response calculation - California, Tech. Rep., February, California Energy Commission, 2003.
[6] J.L. Mathieu, P.N. Price, S. Kiliccote, M.A. Piette, Quantifying changes in building electricity use, with application to demand response, IEEE Trans. Smart Grid 2 (3) (2011) 507–518. http://dx.doi.org/10.1109/TSG.2011.2145010.
[7] K. Coughlin, M.A. Piette, C. Goldman, S. Kiliccote, Statistical analysis of baseline load models for non-residential buildings, Energy Build. 41 (4) (2009) 374–381. http://dx.doi.org/10.1016/j.enbuild.2008.11.002.
[8] R. Weron, Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach, John Wiley & Sons Ltd., 2006.
[9] J.R.M. Hosking, R. Natarajan, S. Ghosh, S. Subramanian, X. Zhang, Short-term forecasting of the daily load curve for residential electricity usage in the Smart Grid, Appl. Stoch. Models Bus. Ind. 29 (6) (2013) 604–620.
[10] J.A. Hausmann, M. Kinnucan, D. McFaddden, A two-level electricity demand model: Evaluation of the connecticut time-of-day pricing test, J. Econometrics 10 (3) (1979) 263–289. http://dx.doi.org/10.1016/0304-4076(79)90085-X.
[11] F.A. Wolak, An experimental comparison of critical peak and hourly pricing: The PowerCentsDC Program, Tech. Rep., Stanford University, Stanford, 2010.
[12] S. Fan, K. Methaprayoon, W.J. Lee, Multiregion load forecasting for system with large geographical area, IEEE Trans. Ind. Appl. 45 (4) (2009) 1452–1459. http://dx.doi.org/10.1109/TIA.2009.2023569.
[13] J.W. Taylor, Triple seasonal methods for short-term electricity demand forecasting, European J. Oper. Res. 204 (1) (2010) 139–152. http://dx.doi.org/10.1016/j.ejor.2009.10.003.
[14] V. Dordonnat, S.J. Koopman, M. Ooms, Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling, Comput. Statist. Data Anal. 56 (11) (2012) 3134–3152. http://dx.doi.org/10.1016/j.csda.2011.04.002.
[15] S. Fan, R.J. Hyndman, Short-term load forecasting based on a semi-parametric additive model, IEEE Trans. Power Syst. 27 (1) (2012) 134–141. http://dx.doi.org/10.1109/TPWRS.2011.2162082.
[16] N. Amjady, Day-ahead price forecasting of electricity markets by a new fuzzy neural network, IEEE Trans. Power Syst. 21 (2) (2006) 887–896. http://dx.doi.org/10.1109/TPWRS.2006.873409.
[17] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: A review and evaluation, IEEE Trans. Power Syst. 16 (1) (2001) 44–55.
[18] T. Hong, Short term electric load forecasting (Ph.D. thesis), North Carolina State University, 2010.
[19] K.X. Perez, W.J. Cole, J.D. Rhodes, A. Ondeck, M. Webber, M. Baldea, T.F. Edgar, Nonintrusive disaggregation of residential air-conditioning loads from sub-hourly smart meter data, Energy Build. 81 (2014) 316–325. http://dx.doi.org/10.1016/j.enbuild.2014.06.031.
[20] D. Benyoucef, P. Klein, T. Bier, Smart Meter with non-intrusive load monitoring for use in Smart Homes, in: 2010 IEEE International Energy Conference and Exhibition, EnergyCon 2010, 2010, pp. 96–101. http://dx.doi.org/10.1109/ENERGYCON.2010.5771810.
[21] F. Farzan, M.a. Jafari, J. Gong, F. Farzan, A. Stryker, A multi-scale adaptive model of residential energy demand, Appl. Energy 150 (2015) 258–273. http://dx.doi.org/10.1016/j.apenergy.2015.04.008.
[22] Y. Ding, S. Pineda, P. Nyeng, J. Østergaard, E.M. Larsen, Q. Wu, Real-time market concept architecture for EcoGrid EU - A prototype for European smart grids, IEEE Trans. Smart Grid 4 (4) (2013) 2006–2016.
[23] G.B. Shrestha, P.a.J. Fonseka, Congestion-driven transmission expansion in competitive power markets, IEEE Trans. Power Syst. 19 (3) (2004) 1658–1665. http://dx.doi.org/10.1109/TPWRS.2004.831701.
[24] O. Corradi, H. Ochsenfeld, H. Madsen, P. Pinson, Controlling electricity consumption by forecasting its response to varying prices, IEEE Trans. Power Syst. 28 (1) (2012) 421–429.
[25] M.T. Hagan, S.M. Behr, The time series approach to short term load forecasting, IEEE Trans. Power Syst. 2 (3) (1987) 785–791. http://dx.doi.org/10.1109/TPWRS.1987.4335210.
[26] S.A.-h. Soliman, A.M. Al-Kandari, Electrical Load Forecasting: Modeling and Model Construction, Elsevier, 2010.
[27] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Vol. 102, Springer, 2006, http://dx.doi.org/10.1016/j.peva.2007.06.006.
[28] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, 2006, http://dx.doi.org/10.1007/978-0-387-40065-5.
[29] A. Martin, J.C. Müller, S. Pokutta, Strict linear prices in non-convex European day-ahead electricity markets, Optim. Methods Softw. 29 (1) (2012) 189–221. http://dx.doi.org/10.1080/10556788.2013.823544, arXiv:1203.4177.
[30] R. Lockhart, R.J. Tibshirani, R. Tibshirani, J. Taylor, A significance test for the lasso, 2014. arXiv:arXiv:1301.7161v1, http://dx.doi.org/10.1214/13-AOS1175.