**Title** | Prediction of Residential Building Demand Response Potential Using Data-Driven Techniques  
---|---  
**Authors(s)** | Kapetanakis, Dimitrios-Stavros; Neu, Olivier; Finn, Donal  
**Publication date** | 2017-08-09  
**Conference details** | The 15th International Building Performance Simulation Association Conference, San Francisco, United States of America, 7-9 August 2017  
**Publisher** | IBPSA  
**Link to online version** | http://buildingsimulation2017.org/; http://www.ibpsa.org/?page_id=962  
**Item record/more information** | http://hdl.handle.net/10197/11545  
**Publisher's version (DOI)** | 10.26868/25222708.2017.439
Prediction of Residential Building Demand Response Potential Using Data-Driven Techniques

Dimitrios-Stavros Kapetanakis¹, Olivier Neu¹, Donal P. Finn¹

¹School of Mechanical and Materials Engineering, University College Dublin (UCD)

Abstract
This paper is concerned with the evaluation of the ability of data-driven predictive models to capture the demand response potential in residential buildings. A mid-floor apartment with an air to water heat pump for space heating, utilised as an archetype dwelling, is simulated using EnergyPlus. The research is focused on forecasting the electrical demand from the heating load for the coldest month of the year, considering two types of DR events, load reduction and load increase. After the generation of the synthetic database, an artificial neural network model and a support vector machine model are examined regarding their ability to predict the electrical demand from heating loads.

Introduction
The integration of renewable energy sources (RESs) requires more flexibility from the power system, due to the variable and uncertain nature of RESs, particularly wind and solar generation. Utilisation of the flexibility offered by electricity demand side management (DSM) and demand response (DR) is one possible strategy. DSM is defined as the modification of customer electricity demand in order to coincidentally effect desirable changes in the electric utility load, in both magnitude and shape, as well as reduce customer expenses (Gellings, 1985). Several advantages are linked with the correct application of DSM, as listed below (Corbin and Henze, 2016; Strbac, 2008; Pina et al., 2012; Warren, 2014):

- reduce generation margin used to handle peak demands;
- improved operational efficiency in generation, transmission and distribution of electric power;
- lower price volatility;
- reduced electricity costs for customers;
- more cost-effective utility system investments, and;
- a cost-effective integration of RESs.

DR is one form of DSM and is defined as the changes in electric usage, implemented directly or indirectly by end-use customers, from their normal consumption patterns in response to certain signals (He et al., 2013). The outcome of DR load shaping alters the electric load profile of electrically driven heating ventilation and air conditioning (HVAC) systems by dissociating in time the demand for electrical and thermal power, which could result in operational benefits on the power system level (Arteconi et al., 2012).

Forecasting of building electrical demand from thermal loads will be of vital importance to grid operators, aggregators and building energy management systems. These forecasts could offer comprehensive knowledge regarding the DR potential in different building types and may be able to offer useful insight to the power system, as already targeted by IEA (Jensen, 2014). Three major categories of models which can simulate or predict the electricity load of individual buildings exist in the literature (Fouquier et al., 2013; Zhao and Magoulès, 2012): building energy simulation, resistance-capacitance network and data-driven models. Predictions of building electrical demand from thermal loads can be estimated using appropriate simulation software (Crawley et al., 2008) when detailed data such as building geometry, occupancy as well as environmental variables are available. In reality, such data are often unknown, especially for older buildings, where uncertainty arising from parameter and occupancy estimation can lead to significant additional modelling efforts (Kwok and Lee, 2011). An alternative way to forecast these loads for residential buildings is to take advantage of smart metering data. These data records include underlying information regarding building thermal response and can be introduced to data-driven models, which utilise extensive assessment of input and output variables, in order to produce accurate predictions (Fouquier et al., 2013). In particular, data-driven models require measured historical data of buildings to accurately predict the electrical demand from thermal loads. An obvious drawback regarding the implementation of data-driven models to capture the DR potential is the fact that it is unlikely to obtain a reliable database of residential buildings that includes historical DR events.

The main objective of the current paper is the evaluation of the performance of data-driven models in capturing DR potential of heating loads in residential buildings. An archetype approach for character-
using the existing building stock in Ireland is adopted in the context of the present research. The implementation of residential building archetypes with the present methodology could provide an estimation of DR potential at city or even national scale.

**Background**

The literature is rich with various models which apply machine learning algorithms to predict the electricity load of individual buildings. The difference of machine learning techniques with building energy simulation models is the fact that they do not require any physical information of the building as input (Foucquier et al., 2013). Heat transfer equations, geometrical parameters and thermal properties are not expected in machine learning algorithms. These algorithms are based on the implementation of a function deduced only from samples or training data, which capture the thermal response of a building (Foucquier et al., 2013). Several algorithms have been used to build predictive models implementing machine learning techniques. The most common algorithms used in the literature to achieve forecasting of building electricity and thermal loads are artificial neural network (ANN) and support vector machine (SVM).

ANNs have been applied to analyse various types of building energy consumption, as well as electrical and heating loads. Kalogirou et al. (1997, 2001) implemented ANN at an early design stage to predict the required heating load of buildings. Input data included the areas of windows, walls and floors, the type of windows and walls, roof classification and the room temperature. The relative error of the network was 3.5%. González and Zamarreño (2005) used an ANN approach to predict the hourly energy consumption in buildings. The inputs of the network were current and forecasted values of ambient air temperature, the current load, the hour and the day. Yang et al. (2005) evaluated the performance of adaptive ANN models that are capable of adapting themselves to unexpected pattern changes in the incoming data. Two adaptive models were proposed and evaluated, accumulative training and sliding window training. These models can be used for real-time on-line building energy prediction. Moreover, they used both simulated (synthetic) and measured datasets. Neto and Fiorelli (2008) compared an ANN model with a building energy simulation model, developed with Energy-Plus, regarding their ability to forecast the energy demand in an administration building in Sao Paulo, Brazil. Two ANN models were developed, one with five input variables (external temperature, humidity, two solar radiation parameters and day-type) and one with two input variables (external and internal temperature). Results from both ANN and EnergyPlus models appeared to be close indicating their suitability for energy consumption forecast. Ekici and Aksoy (2009) used an ANN to predict building energy needs benefiting from orientation, insulation thickness and transparency ratio. A back propagation network was preferred and available data were normalised before being presented to the network. The calculated values compared to the outputs of the network gave satisfactory results with a deviation of 3.4%. More recently, Burger and Moura (2015) formulated an ensemble machine learning method that performs model validation and selection in real time using a gating function. The ensemble models was designed to forecast building electricity demand, by learning from electricity demand data streams, while requiring little knowledge of energy end-use. The models was tested both on commercial and residential buildings. In particular, 24 residential buildings were used and the generated electricity demand forecasts had a mean absolute percent error of 55.8%.

SVM models have been used more recently for predicting energy consumption in buildings. Lai et al. (2008) employed the SVM as a data mining tool for the prediction of the electrical consumption in residential sector in the region of Tohoku, Japan. Data from outdoor and indoor air temperatures and humidities were considered as input parameters. Kavaklioglu (2011) used a regression SVM model to predict the electricity consumption in Turkey. Electricity consumption was predicted until 2026 using data from 1975 to 2006. The radial basis was used as the kernel function of the SVM model, while the input variables were socio-economic indicators such as population, Gross National Product, imports and exports. Results illustrated that electricity consumption can be modelled using a regression SVM model, which can be used to predict future electricity consumption. Che et al. (2012) developed an adaptive fuzzy combination model based on a self-organizing map, an SVM and the fuzzy inference method to predict the electrical load in New South Wales. It was demonstrated that the adaptive fuzzy combination model can effectively count for electric load forecasting with good accuracy and interpretability at the same time. Li et al. (2010) predicted the annual energy consumption of residential buildings using four different machine learning algorithms such as SVM and three types of ANN, the traditional back propagation neural network, the radial basis function neural network as well as the general regression neural network. Their study was based on 59 residential buildings in Guangdong, China and the obtained results demonstrated that the SVM model was more accurate than the ANN models.

The research field related to building energy consumption (both electrical and thermal loads) forecasting has been very active, involving the implementation of various ANN and SVM data-driven models. Nevertheless, it is clear from the literature that little attention has been given to the ability of data-driven models to capture the DR potential of resi-
dential buildings. The reason behind this research gap might be the fact that it is unlikely to obtain a reliable database of residential buildings that includes historical DR events.

### Methodology

To evaluate the ability of data-driven models to capture the DR potential of residential buildings, a synthetic database for a reference dwelling is generated utilising EnergyPlus. In this way the deficiency of obtaining a reliable database of residential buildings that includes historical DR events is avoided. The methodology developed and implemented in this paper consists of four stages:

- **i** Generation of a synthetic database, for an archetype dwelling;
- **ii** Introduction of DR events at the synthetic database;
- **iii** Development of data-driven predictive models, and;
- **iv** Evaluation of the accuracy of predictive models.

#### Generation of Synthetic Database

The reference dwelling used in this paper, is an archetype of a mid-floor apartment in Ireland, as illustrated in Figure 1 (Neu et al., 2014). In general, the development of building archetype models, being representative of a group of dwellings and dwelling loads, allows modelling and simulation of the performance of building stock as a whole. Moreover, this approach complements a power system perspective on the aggregated DR potential offered by residential dwellings through the implementation of DSM strategies, as emphasized by Ma et al. (2013).

The mid-floor apartment utilised is considered an archetype due to the integration of the necessary operational data with high space and time resolution (15-minutes). This data subset is built upon the bottom-up approach proposed by Neu et al. (2013), by implementing an adjusted Markov-Chain Monte Carlo technique, pioneered by Richardson et al. (2008). This technique is applied to the 2005 Irish time-use survey (TUS) activity data (ESRI, 2005a), thus taking into account end-user behaviour. Activity-specific profiles for occupancy, electrical appliance and lighting use, domestic hot water (DHW) demand (Neu et al., 2013, 2014, 2016) are developed following this technique. These operational inputs were verified against surveyed (ESRI, 2005b) and metered data (CER, 2012).

Moreover, the simulated outputs from the dwelling archetype model were validated against the dwelling energy assessment procedure (DEAP) methodology (SEAI, 2012), in terms of annual electricity, space and water heating requirements and daily DHW demand (Neu et al., 2014, 2016). The validation was performed using the International Weather for Energy Calculations (IWEC) data (ASHRAE, 2001). In addition, the archetype is considered over two different construction periods, new and existing. The new construction of the archetype was built in accordance with the latest Irish building regulations (DECLG, 2011), while existing constructions do not meet the standards set by these regulations, particularly in terms of the building envelope insulation level, infiltration and ventilation levels. This approach allows variations in energy performance and potential electrical flexibility resource across different construction periods to be captured on a 15-minute basis.

The system used for space heating of the archetype is an air to water heat pump. In the context of investigating the DR potential of residential buildings, space and water heating systems are expected to offer larger potential for flexible operation than electrical appliances and non-heating loads (CER and UR, 2011). Furthermore, the shift from fossil fuel-based space and water heating systems, such as gas or oil-fired boilers, towards electrified and more energy-efficient technologies, such as heat pumps, is foreseen both globally (IEA, 2011) and in the European Union (Eurelectric, 2011). Among those technologies considered by the IEA (2011) for space and water heating, heat pumps contribute to more than 40% of building-related carbon dioxide emissions by 2050. Thus, a heat pump is chosen in this study as a replacement for gas and oil-fired boilers to meet the space and water heat demands of the new and existing archetypes.

#### Introduction of DR Events at the Synthetic Database

To evaluate the ability of data-driven predictive models to capture the DR potential of the electrical demand from the heating load, the coldest month of the year was selected to be examined. The location of the archetype was to be in Dublin, Ireland, hence the coldest month based on IWEC data is February. The archetype dwellings, both new and existing, are initially simulated without the presence of any DR event and subsequently with a DR event occurring daily and weekly. The two scenarios investigated in the current research, weekly and daily, reflect a moderate and an excessive implementation of DR event strategy from grid operators, respectively.

Furthermore, two types of DR events, load reduction and load increase, by changing the setpoints of the heating system thermostat, are considered, while maintaining the thermal comfort of the occupants.

![Figure 1: Mid-floor apartment archetype.](image-url)
within acceptable limits. More precisely, in response to a signal sent to the thermostatically controlled space heating system, the temperature setpoints are varied. These are either decreased or increased in value until their minimum or maximum limits of operation, thus simulating a load reduction or increase event, respectively. A one-hour load variation (reduction or increase) event is considered, lying within the range of durations specified by EirGrid (2015) for demand side resources, which varies from half an hour to two hours. Following the one-hour DR event (reduction or increase), the setpoint values are returned to their initial settings.

The occurrence time of the load reduction DR event is selected to be from 18:00 to 19:00, while regarding the load increase DR event is set from 16:00 to 17:00, for both scenarios under investigation, weekly and daily. The load reduction DR event is initiated at 18:00, in order to curtail the evening peak in consumption of residential buildings. Likewise, the load increase DR event is designed to avoid the evening peak, by pre-heating the archetype starting at 16:00. Moreover, the days when the DR event is occurring for the weekly scenario are selected randomly and are Monday for the first week, Thursday for the second week, Wednesday for the third week, Friday for the fourth week and Tuesday for the fifth week.

Development of Data-Driven Models
Following the generation of the synthetic database, ANN and SVM data-driven models are examined regarding their ability to predict the electrical demand from heating loads, with and without DR events. The development of the models is performed considering only the weekdays of the coldest month. Weekend days are excluded due to the completely different electrical demand profiles of residential buildings over weekends.

Initially, the datasets are divided into two partitions, training and testing, which are used to train and test the developed predictive models. Each partitions consists of eleven weekdays. In particular, the training partition consists of the weekdays from 30th of January to 13th of February, while the testing partition consists of the weekdays from 14th to 28th of February. The input variables of the data-driven predictive models in the context of this research are time, ambient air temperature, ambient air relative humidity, solar radiation, wind speed, zone air temperature of the four zones of the archetype dwelling and a binary variable indicating if there is a DR event in place or not. The output of the predictive models is the electrical demand associated with the heating loads of the archetype. Moreover, individual predictive models are developed for predicting the electrical demand from heating loads, when daily and weekly DR events are included in the dataset.

The ANN predictive models are ensemble models developed using boosting, which generates a sequence of models to obtain more accurate predictions. Boosting produces a succession of models, each of which is built on the same training partition of the dataset. Prior to building each successive model, the input variable measurements are weighted, based on the residuals of the previous model. Measurements with large residuals are given relatively higher analysis weights, so that the next model focuses on predicting these records better. In addition, the ANN models connect the input to the output variable through the hidden layers using the multilayer perceptron (MLP) structure and the sigmoid activation function.

The SVM predictive models use a machine learning regression algorithm that maximizes the prediction accuracy without overfitting the training data. A kernel function is implemented when applying the SVM algorithm in order to overcome the presence of non-linearity relationships between input and output variables. The kernel function selected is the polynomial one.

The IBM SPSS Modeler 14.2 software (IBM Corp., 2011) was used for the development of the predictive models using a computer with an Intel Core i7-3630QM processor and 8 GB of DDR3 RAM. The settings of the ANN and the SVM models are selected automatically from the software with the objective to enhance the models accuracy. The ANN model structure can not be extracted from the software when the boosting option is selected, while the details of the most accurate SVM algorithm are as follows:

- regularization parameter (C) equal to 10;
- regression precision epsilon (\(\epsilon\)) equal to 0.1;
- gamma (\(\gamma\)) equal to 1;
- bias parameter equal to 0.8, and;
- degree of complexity set to 6.

Evaluation of Accuracy of Predictive Models
The evaluation of the accuracy of the data-driven predictive models is based on their performance with the testing partition of the synthetic datasets. The overall accuracy of each predictive model is calculated based on the coefficient of variation of the root mean square error (CV-RMSE):

\[
CV - RMSE = \frac{RMSE}{\bar{y}} = \sqrt{\frac{\sum_{i=1}^{n} |y_{i} - \hat{y}_{i}|^2}{n}} (1)
\]

where, \(y\) are the actual values, \(\hat{y}\) are the predicted values of the electrical demand associated with the heating loads of the archetype, \(\bar{y}\) is the mean of the actual values and \(n\) is the total number of timesteps summed up at the testing partition period.

Further to the overall accuracy of the predictive models, specific days of the testing dataset are monitored in order to evaluate the ability of the predictive models to capture the occurring DR events and sudden changes to the electrical demand from heating loads of the archetype.
Table 1: CV-RMSE of data-driven models.

<table>
<thead>
<tr>
<th>Building Construction</th>
<th>Scenario</th>
<th>ANN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing</td>
<td>Without DR event</td>
<td>2.448</td>
<td>2.028</td>
</tr>
<tr>
<td></td>
<td>Daily DR load increase</td>
<td>1.788</td>
<td>1.551</td>
</tr>
<tr>
<td></td>
<td>Daily DR load reduction</td>
<td>2.850</td>
<td>1.961</td>
</tr>
<tr>
<td></td>
<td>Weekly DR load increase</td>
<td>3.192</td>
<td>1.916</td>
</tr>
<tr>
<td></td>
<td>Weekly DR load reduction</td>
<td>2.853</td>
<td>1.996</td>
</tr>
<tr>
<td>New</td>
<td>Without DR event</td>
<td>2.157</td>
<td>2.102</td>
</tr>
<tr>
<td></td>
<td>Daily DR load increase</td>
<td>1.391</td>
<td>1.342</td>
</tr>
<tr>
<td></td>
<td>Daily DR load reduction</td>
<td>2.092</td>
<td>1.813</td>
</tr>
<tr>
<td></td>
<td>Weekly DR load increase</td>
<td>2.619</td>
<td>1.788</td>
</tr>
<tr>
<td></td>
<td>Weekly DR load reduction</td>
<td>2.397</td>
<td>1.925</td>
</tr>
</tbody>
</table>

Results and Discussion

Once the synthetic database is generated using the archetype dwelling, ANN and SVM data-driven models are investigated regarding their ability to predict the electrical demand from heating loads, with and without DR events. The overall results regarding the accuracy of the data-driven models are summarised in Table 1. It is observed that for all the scenarios examined, the SVM predictive model performs better than the ANN model. Furthermore, it is noticed that the most accurate results, of both construction periods, are obtained for the scenario when a load increase DR event is occurring daily. The least accurate results for the ANN predictive models, of both construction periods, are captured for the scenario when a load increase DR event is occurring weekly. In general, the SVM predictive models forecast the electrical demand from heating loads at the same level of accuracy, while the least accurate results, of both construction periods, appear for the scenario without the presence of a DR event.

Additionally, to the overall accuracy of the predictive models, specific days of the testing partition of the dataset are plotted, in order to visualise and evaluate the ability of the predictive models to capture the occurring DR events as well as other sudden changes to the electrical demand from heating loads of the archetype. The days selected for each scenario are the 21\textsuperscript{st}, 20\textsuperscript{th} and 28\textsuperscript{th} of February, when none, daily and weekly DR event are included in the dataset, respectively.

The performance of the ANN and SVM predictive models for the scenario without DR event, is presented in Figure 2 (a) and (b) for the existing and new construction period, respectively. The visualisation of the prediction of the data-driven models provides the ability to make useful observations. Initially, it is noted that neither the ANN or the SVM models, where able to predict the sudden changes of the electrical demand from heating loads appearing

![Figure 2: Prediction of electrical demand from heating load without DR event. (continued)](image-url)
within the day, both for the existing and the new construction of the archetype. These sudden changes were observed with various magnitudes and occurrence times from one day to another, thus making it difficult for the predictive models to capture them accurately. Moreover, the ANN predictive model did not manage to accurately forecast the electrical demand, while generated some negative predictions as well. On the contrary, the SVM predictive model illustrates the ability to forecast accurately the base electrical demand of the existing and new construction archetype.

Figure 3 (a) and (b), depicts the performance of the predictive models for the scenario with daily DR load increase event, for the existing and new construction period, respectively. The load increase DR event is set from 16:00 to 17:00 and it is noticed that both ANN and SVM models manage to capture that increase. Overall, once again it is observed that the SVM model manages to forecast the base electrical demand of the existing and new construction archetype.

Figure 3: Prediction of electrical demand from heating load with daily DR load increase event. (continued)
demand better than the ANN model. Regarding the archetype with the existing construction, two peaks are taking place within the day under examination, the first of which none of the models manages to capture, while the second one is captured only by the ANN model. Despite the fact that the ANN model predicted the second peak, its predictions prior to the peak are far from the actual electrical demand. There is only one sudden peak for the archetype with the new construction, but none of the models was able to capture it.

The performance of the predictive models for the scenario with daily DR load reduction event, is displayed in Figure 4 (a) and (b) for the existing and new construction period, respectively. The load reduction DR event is set from 18:00 to 19:00 and it is noticed that both ANN and SVM models manage to capture that reduction for the archetype with the existing construction type. On the contrary, only the SVM model performed closely to the actual electrical demand.
demand during the DR event period for the new construction archetype. An interesting observation for this scenario, is that none of the predictive models manage to capture the rebound effect occurring at 19:00 right after the DR event ended, for both existing and new construction types of the archetype. Furthermore, similarly with the previous examined days, the SVM model manages to forecast the base electrical demand better than the ANN model, while none of them captures the sudden peak occurring within the day under examination.

Figure 5 (a) and (b), presents the performance of the predictive models for the scenario with weekly DR load increase event, for the existing and new construction period, respectively. During the load increase DR event (from 16:00 to 17:00) both ANN and SVM models overestimate the potential increase for the new construction type of the archetype. Regarding the archetype with the existing construction, the ANN slightly underestimates and the SVM model

Figure 4: Prediction of electrical demand from heating load with daily DR load reduction event.

Figure 5: Prediction of electrical demand from heating load with weekly DR load increase event. (continued)
overestimates the potential increase. Again it is observed that for both construction types the SVM model manages to forecast the base electrical demand better than the ANN model and that none of the models captured the sudden peaks of the electrical demand.

Finally, the performance of the predictive models for the scenario with weekly DR reduction event, is given in Figure 6 (a) and (b) for the existing and new construction period, respectively. It is observed that during the load reduction DR event (from 18:00 to 19:00), for the existing construction type, the SVM model performs quite well, while the ANN model does not manage to capture the reduction. The results from the new construction illustrate that the ANN model overestimates the potential reduction that can be achieved, while the SVM model does not capture the reduction. The obtained results of this scenario suggest that training the models with weekly events might not be sufficient for accurate predictions.
Conclusions
To summarise, it is shown that the predictive models are capable of forecasting accurately the base electrical demand of a residential archetype, as it is well documented in the literature, but at the same time are not able to generate accurate predictions when sudden peaks appear, which is due to the high variability of the residential electrical demand. Moreover, regarding the ability of the predictive models to capture DR events, results indicate that SVM models can achieve higher accuracy in predictions, while capturing better the DR events compared to the ANN models. In general, it is noticed that the models managed to handle better DR events that target a load increase rather than a load reduction. The issues related to the load reduction DR events, for the scenarios that occur daily and weekly, are the inability to capture the rebound effect and the over/under estimation of the reduction potential, respectively. Future research work includes the application of the methodology presented in this paper to different dwelling archetypes. Additionally, other types of heating systems will be under investigation to evaluate if more accurate predictions can be generated. In this way, the effectiveness of ANN and SVM model in capturing DR potential of heating loads in other types of residential buildings, with different heating systems, will be examined. Finally, through this future research work, the reasoning for the performance of the predictive models as well as the generalisation of the obtained results will be assessed.

References


