LBNL 0000

Lawrence Berkeley National Laboratory

Modeling Study on Flexible Load's Demand Response Potentials for Providing Ancillary Services at the Substation Level

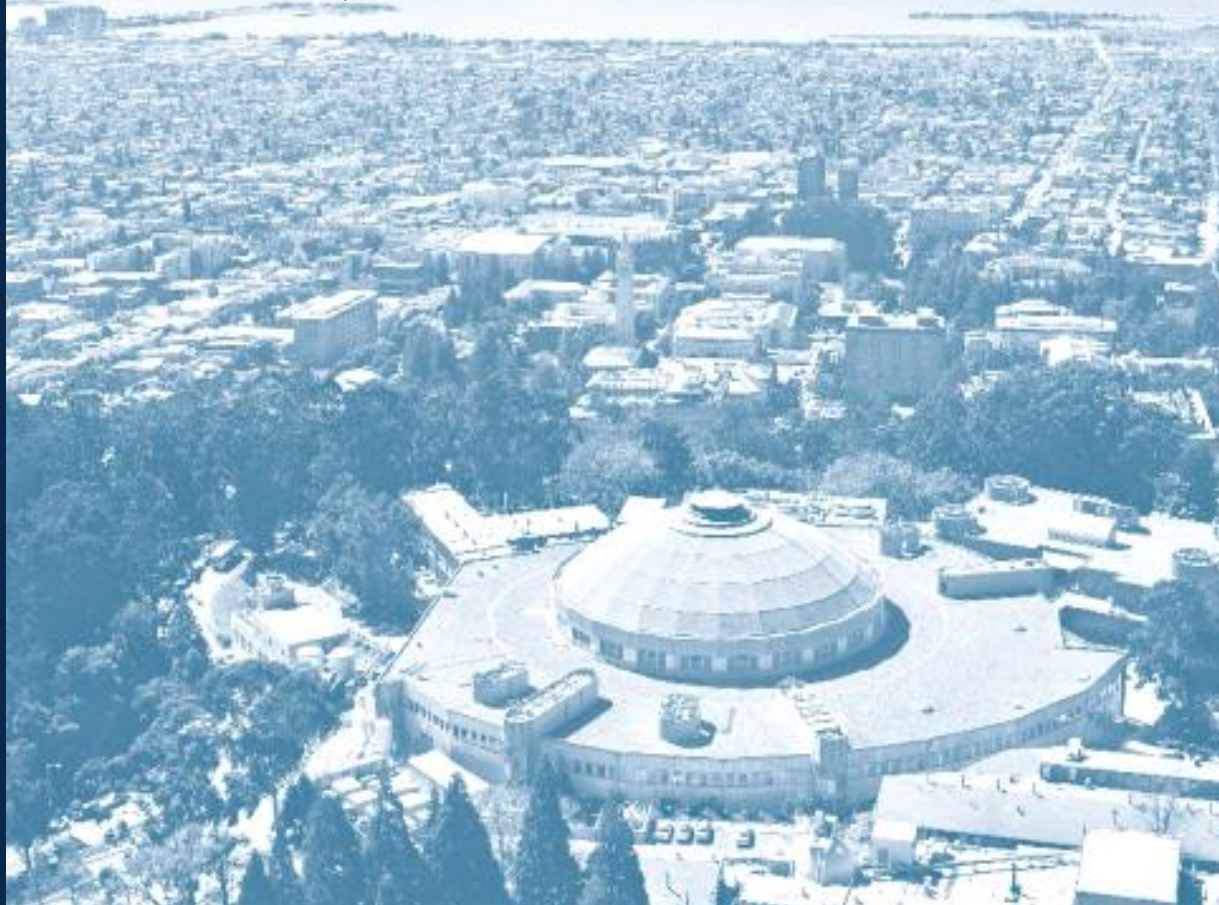
Ryan Tulabing[1], Rongxin Yin[1], Nicholas DeForest[1], Yaping Li[2], Ke Wang[2], Taiyou Yong[2], Michael Stadler[1]

[1]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[2]China Electric Power Research Institute, China

To be Published in the Journal of Electric Power Systems Research

June 2016

The work described in this study was performed and coordinated by the Grid Integration Group of Lawrence Berkeley National Laboratory and supported by the State Grid Corporation of China Project (DZN17201300197, Study on Key Technologies for Power and Frequency Control of System with Source-Grid-Load Interactions).



Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Modeling Study on Flexible Load's Demand Response Potentials for Providing Ancillary Services at the Substation Level

Ryan Tulabing[a], Rongxin Yin[a], Nicholas DeForest[a], Yaping Li[b], Ke Wang[b], Taiyou Yong[b], Michael Stadler[a,*]

[a]*Energy Storage and Distributed Resources Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*
[b]*China Electric Power Research Institute, China*

Abstract

Demand Response (DR) is an important component for the establishment of smart electricity grids. It can decrease the system peaks through load shedding or shifting and optimize the utilization of the existing grid assets, which delays the need for costly upgrades. DR can also enable the integration of intermittent and distributed energy resources (DER) into the existing electricity grid. Fast DR from aggregated flexible loads can provide ancillary services (AS) to absorb grid disruptions and may replace the expensive fast-ramping reserve generation units. This study presents a methodology for load aggregation based on the prioritization of loads according to their flexibility. Different flexible load types are categorized as Thermostatically Controlled Loads (TCL), Urgent Non-TCL, Non-urgent Non-TCL, and Battery-based Loads. Models based on their physical behaviour are developed and simulations performed to apply the proposed aggregation and control algorithm. Results show that the loads during peak hours can be shed off without rebound demand spikes after the DR event commonly seen in other types of DR programs. The algorithm also automatically adjusts the power demand according to the output of the distributed renewable generation, mitigating disruptions due to variations of the DER output. Additionally, the algorithm is able to adjust the load demand dynamically according to the fluctuations of electricity price.

Keywords: demand response, thermostatically controlled loads, regression models, two-state model, simplified DR potential estimation.

1. Introduction

Grid frequency control and power balancing are traditionally done by ancillary units commonly composed of fast ramping generation units such as gas turbines and diesel generators. However, as new rules to reduce greenhouse gas emissions are implemented, system operators are turning to non-fossil fuel powered resources. Increased penetration of renewable generation from solar and wind, which are intermittent and non-dispatchable, is further driving the need for fast ramping resources. From this point of view, the use of flexibility of demand-side resources and availability of real-time signals communication in the electricity grid enables the interactions between the supply and the demand.

Over the past decades, considering the grid issues of power imbalance and peak demand, Demand response (DR) has proven to be a viable option by load shedding and load shifting in response to the need of grid. A number of studies have demonstrated the traditional DR for emergency load relief, peak load management, and price responsive demand [1, 2, 3, 4]. However, the use of DR for ancillary service is different from that of traditional DR applications as DR for AS require fast response and high accuracy. Recent studies have demonstrated the use of demand-side resources to provide ancillary service in the electricity market[4, 5, 6]. AS can be classified into three categories: Regulation, Flexibility, and Contingency [7]. Ma

*Corresponding author: Michael Stadler, mstadler@lbl.gov

et al. [5] and Kirby [8] described generalized DR product definitions for load participation in AS, energy, and capacity markets. Those DR product are defined by the response time, the length of the response, the time to fully respond and the event times being called. Regulation service refers to the capacity to respond to random deviations from the scheduled net load. Response time for this type of ancillary service vary between 30 seconds to 5 minutes and usually lasts for 15 minutes. Flexibility ancillary service addresses the large unforeseen deviation of wind or solar output responding as fast as 5 minutes to 20 minutes for a duration of 1 hour. Meanwhile, contingency service is allocated for immediate response to sudden loss in generation. Contingency services are required to respond as fast as 1 minute to 10 minutes holding for a duration of 30 minutes or less [7]. In the US electricity market, such as CAISO (California ISO) and PJM region, frequency service requires 4 seconds response to track automatic generation (AGC) control signals.

From the perspective of DR enablement, DR can be considered broadly to fall into two categories: direct control and indirect control. Direct control enables the grid operators to directly turn on or off the customer's loads or change the operating setpoints (e.g. thermostat control) after a short notice. Alternatively, under indirect control, grid operators send requests to reduce load demand to customers, who have the choice to participate or not. In the context of AS described above, demand response has to be fast and automatic (i.e. Auto-DR) [9]. The variability in customer response time makes indirect control less reliable compared to the direct control DR, and unsuited for this application. Different control methodologies for harnessing the DR potential of Heating Ventilation and Air Conditioning (HVAC) system were evaluated by [10, 11]. Results show that significant demand relief can be taken from temperature reset control and pre-cooling control but with energy and cost penalties. Highest demand relief can be attained by curtailment control but can only be used in shorter duration as the indoor temperature quickly exceeds the comfort level for most humans. For demand response lasting for several hours, pre-cooling control is among the best options [12, 13, 14]. However, with regards to DR for AS, battery-based storage and electric vehicle (EV) have proven to be viable options for grid application, ancillary services such as frequency regulation in particular [15, 16, 17, 18]. When considered as flexible resources, a number of recent studies have demonstrated the use of EVs for increasing penetration of renewable generation resources. In [19], the authors design three a suitable modeling of electric vehicles with three types of controls (night charge, intelligent charge and vehicle to grid) to analyze the impact of EVs on energy systems. The authors of [16] conduct several simulations to show the adoption of advanced centralized EV charging control strategies and allow the integration of a larger number of EVs in the system. On the other hand, the adoption of a local level of control will allow a better operation performance of increasing penetration of intermittent and variable renewable generation resources installed in isolated power systems. In [20], the authors investigate the use of plug-in EVs (PEVs) to balance the fluctuation of renewable energy sources and study the benefits of fleet EVs in response dynamic price signals. In [21], the authors explore the potential of PEVs to balance variability and uncertainty from wind and solar generation resources. The authors demonstrate the use of a large number of PEVs to provide ancillary services in the regulation (secondary frequency control) time frame by leveraging emerging Information and Communication Technologies, and conclude that roughly 3 million PEVs with a charging rate of 3.3 kW and no V2G capability would suffice to supply a large part of the regulation up and down demand in California.

In the domain of demand response in buildings, recent studies show that using commercial and residential HVAC load control in grid operations can provide power regulation and ancillary services [22, 23, 24, 25, 26, 27, 28]. Most research have demonstrated the value of demand-side flexible loads individually at each category of DR product. However, at the substation level, it is quite challenging to aggregate different type of flexible loads at the same order due to different response characteristics (e.g. response availability, depth and duration). There is a need to provide a solution to aggregate, control and optimize each type of flexible load's DR resource in the grid operation.

In this study, we develop a suite of bottom-up physical models to quantify aggregate DR potential from residential sector. Specifically, we propose a general methodology for a DR controller which aggregates common types of residential electrical loads, EV and storage, and predicts their potential to provide demand response resource. The paper is organized as follows: Section 2 describes the methodology for modelling each type of flexible load in residential sector, as well as models for distributed energy resources (DER) like solar and wind. Section 3 introduces the definition of flexibility for each load type, and the power allocation

algorithm that prioritizes loads based on their flexibility. Section 4 uses the model detailed in Section 2 and control algorithms detailed in Section 3 to demonstrate the use of aggregated DR resource under several scenarios. Finally, the conclusion and recommendations for future investigations are presented in Section 5.

2. Methodology

This study uses the bottom-up approach in modeling customer loads and their corresponding DR potential. Individual models of electrical loads are developed based on their physical properties and behavior. The models are then aggregated in a simulation to mimic the load demand at the substation level where the proposed controller should be located. Figure 1 illustrates the setup of the controller. Loads will send requests to turn on, then the controller will prioritize all requests and grant permission according to available power from the DER and the grid. It is assumed that the communication line between the aggregator and the loads exists and that the central controller has the ability to prohibit a load from starting. Two conditions are required for a load to start: load request trigger, and permission from the central controller. The central controller gathers the load status or simulate the load status, prioritize the loads according to their flexibility, then allocate power by sending a "permission to start" signal to the loads. If a particular load requests to start but no permission has yet been granted from the central controller, it shall wait until the permission to turn on is given – inherently shifting the load demand.

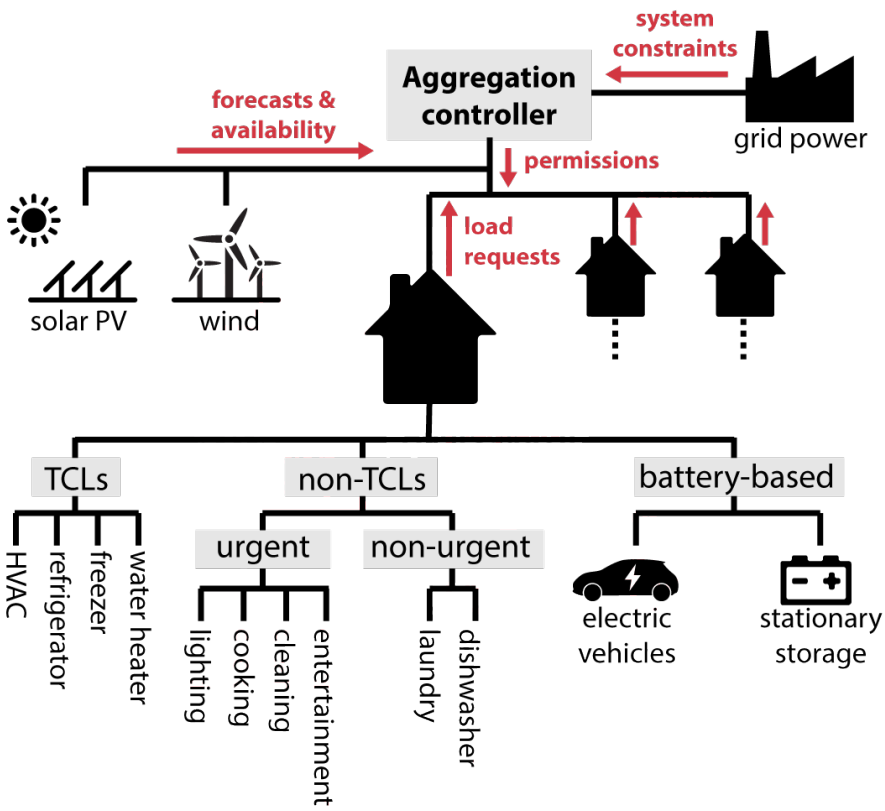


Figure 1: Aggregation of flexible loads

2.1. Load Classification

Electrical loads can be classified as: Thermostatically Controlled Loads (TCL); Non-thermostatically Controlled Loads (Non-TCL); and Battery-based Loads. TCLs include HVACs, water heaters, refrigerators,

and freezers. Non-TCL loads can further be classified as urgent or non-urgent. Urgent loads are types of load that has to respond to user's request instantaneously after the switch is turned on (e.g. lights, cooking appliances, entertainment appliances). Non-urgent loads are those that can be started after some allowable time delay such as dishwashers, washing machines, and clothes dryers. Battery-based loads, including electric vehicles (EV) and stationary batteries, store energy and should be fully charged in a specified time as required by the user. Battery-based loads are assumed to be interruptible and can be delayed as long as they meet their required charge status on the time. By inference, TCLs and battery-based loads have the capacity to store energy (i.e. thermal or chemical) hence, they are inherently flexible. Non-urgent non-TCL loads are also flexible as their starting time can be delayed.

2.2. Load Characterization

2.2.1. TCL Model

The most common method to model TCLs is the Equivalent Thermal Parameter (ETP) model. A schematic representation of this modeling approach [29] is shown in Figure 2.

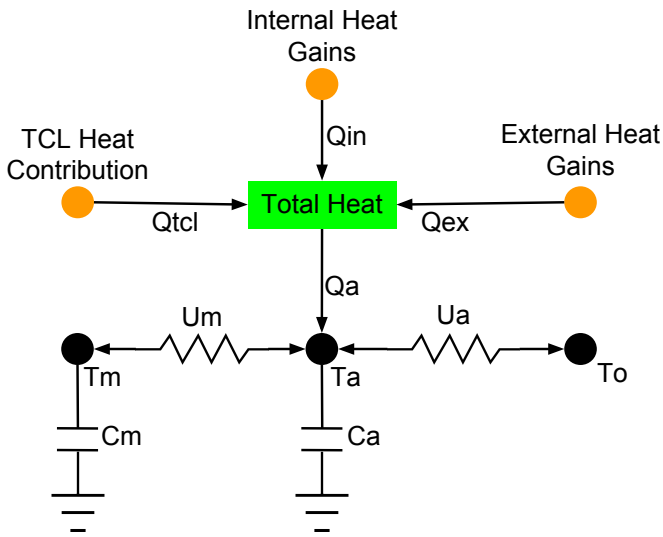


Figure 2: Representation of Equivalent Thermal Parameter (ETP) model

In this representation, U_a is the coefficient of heat transfer between the inside air and the outside environment, U_m is the coefficient of heat transfer between the mass inside the thermal zone and the inside air, C_a is the heat capacity of air, C_m is the heat capacity of the internal mass, Q_a is the net heat introduced to the inside air, T_a, T_o and T_m are temperatures of the inside air, outside environment, and the internal mass respectively. Q_a is the sum of the external heat gain (Q_{ex}), internal heat gain (Q_{in}), and the heat gain/loss due to the TCL (Q_{TCL}).

The ETP model is described mathematically by the following state space function [29]:

$$\dot{x} = Ax + Bu \qquad\qquad y = Cx + Du \tag{1}$$

$$\dot{x} = \begin{bmatrix} \dot{T_a} \\ \dot{T_m} \end{bmatrix} \qquad\qquad x = \begin{bmatrix} T_a \\ T_m \end{bmatrix} \qquad\qquad u = 1 \tag{2}$$

$$A = \begin{bmatrix} -(\frac{U_m}{C_a} + \frac{U_a}{C_a}) & \frac{U_m}{C_a} \\ \frac{U_m}{C_m} & \frac{-U_m}{C_m} \end{bmatrix} \qquad B = \begin{bmatrix} \frac{T_o U_a}{C_a} + \frac{Q_a}{C_a} \\ 0 \end{bmatrix} \tag{3}$$

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4}$$

To account the heat gains or losses due to ventilation and leakages, the following term is added in the differential function for T_a:

$$\dot{m} C_p (T_o - T_a) \tag{5}$$

where, $\dot{m}$ is the mass flow rate, and C_p is the specific heat capacity of air or water (in the case of a water heater).

The common control algorithm for TCLs based on the temperature set-point and the tolerance (dead-band) is used in this study. However, an additional decision gate has been established as prerequisite before a particular TCL turns on.

2.2.2. Battery-based Load Model

The usual charging process of battery-based loads starts with a constant current (CC) mode followed by a constant voltage (CV) mode [30]. At lower charge level, the charging current is fixed at a certain value, based on the charger capacity while the voltage is allowed to swing. When a certain charge level has been reached, the charging voltage is fixed at a certain value while the current is allowed to fluctuate. In this study, it is assumed that the power demand (voltage x current) for charging stays constant while the ratio between voltage and current changes.

Given the capacity of a battery in kilo-Watt-hours (kWh), the length of time for charging (or discharging) is a function of the charging power.

$$t_f - t = \frac{C_B(1 - SOC_0)}{P_{charging} \eta_{charging}} \tag{6}$$

where, t_f is the predicted time when the charging finishes [hour], t is the current time [hour], C_B is the capacity of the battery (kWh), SOC_0 is the existing state of charge [%], $P_{charging}$ is the charging power [kW], and $\eta_{charging}$ is the charging efficiency.

The SOC can also be determined as a function of charging power and the charging time .

$$SOC = SOC_0 + \frac{P_{charging} * \Delta t * \eta_{charging}}{C_B} \tag{7}$$

where, Δt is the duration of charging ($\Delta t = t_f - t$).

2.2.3. Urgent Non-TCL Model

Urgent Non-TCL (UNTCL) loads include lighting, and miscellaneous appliance (e.g. for entertainment, cooking, cleaning, etc.) that directly interacts with the end-users. The models used to account the demand variations from these type of loads are based on a regression curve that fits the typical residential load curve from a suite of prototype buildings. The derived regression functions depict the occupancy schedule on the time of the day. Since these loads need to start instantaneously, their flexibility is zero and they are prioritized during the power allocation process. The following equation expresses the total power demand of the urgent non-TCLs (P_{UNTCL}).

$$P_{UNTCL}(t) = P_{light}(t) + P_{misc}(t) \tag{8}$$

where, P_{light} and P_{misc} are power demand from lighting and miscellaneous appliance loads with hourly operational schedules.

2.2.4. Non-urgent Non-TCL Model

Some loads are not required to start instantaneously if allowed by the end-user. For load such as a dishwasher or clothes drier, a user may not be sensitive to the start time of the load, but is likely to be sensitive to the finish time. Non-urgent non-TCL loads can be characterized according to the earliest time it can start (t_{start}), duration of its operation (Δt), power demand during the operation (P_{rated}), and the latest time it needs to finish its task (t_{finish}) [31].

$$P_{NNTCL} = f\left(P_{rated}, \Delta t, t_{start}, t_{finish}\right) \tag{9}$$

2.3. Power Supply Limitation

In the traditional set-up of controlling the balance between the electricity supply and demand, the generators have to adjust simultaneously and follow the variability of the demand. However, this is not applicable to DER which rely on uncontrollable inputs such as wind speed and solar radiation. A shift from this paradigm is therefore necessary – load demand should follow or adjust according the available power supply. To demonstrate the capability of flexible loads to absorb the grid disruptions caused by electrical loads and DERs, this study proposes a maximum power limit that the loads can consume at a particular time.

$$P_{limit} = P_{base} + P_{DER} \tag{10}$$

If the base power (P_{base}) produced by traditional generators is allowed to be constant, any fluctuations in the output of DER (P_{DER}) would be reflected in the load demand limit (P_{limit}). Consequently, at instances when DER output is high, more load demand requests shall be granted to consume electricity. Conversely during low DER production, the load consumption would automatically adjust and delay the start-up of some flexible loads.

Electricity price can also be included as an incentive to reduce demand. As the electricity market price increase, demand from flexible loads should fall in response. Under this approach the market price factor will only be applied to the base generation and not to DER which generally operate at lower costs. Thus, equation 10 can be modified to:

$$P_{limit} = P_{base}K_p + P_{DER} \tag{11}$$

where, K_p is the price factor defined as:

$$K_p = \frac{A * \Phi_{limit}}{\Phi_{market}} \tag{12}$$

where, A is a predetermined constant, Φ_{limit} is the price threshold, and Φ_{market} is the market electricity price.

Another approach is to define a total cost limit based on the price threshold and the allocated power.

$$C_{limit} = P_{allocated} * \Phi_{limit} \tag{13}$$

C_{limit} is the highest expected cost [$] i.e. when the price of the electricity to be purchased is Φ_{limit}. Upon power allocation, the sum of electricity cost (i.e. demand x price) should not exceed the limit C_{limit}. Thus, if current electricity price is beyond the price threshold, only the loads of higher priority (i.e. lower flexibility) shall be served while the loads of lower priority will be shed off or shifted. This approach has been adopted in the simulation stage of this study.

The management of loads during periods of low DER production high demand, or high market prices is done automatically by the central controller. Loads are prioritized so that some loads with more flexibility will be delayed while loads of higher priority are served first. The prioritization algorithm for different loads is discussed in details in the next section. Meanwhile, the models used to forecast DER output in this study are discussed in the following sub-section.

2.3.1. Solar Power Model

The output of the distributed solar PV (i.e. behind-the-meter PV), which are usually installed on the rooftop, are calculated as:

$$P_{solar} = \frac{P_{installed} * G_T * \eta_{inverter}}{G_{standard}} \tag{14}$$

Where, P_{solar} is the predicted power output from the solar panels [Watts], $P_{installed}$ is the installed capacity of behind-the-meter PV panels, G_T is the global solar irradiance received by the solar panels [Watts/m2], $G_{standard}$ is the standard Irradiance value used by PV manufacturers to test their products (i.e. 1000 Watts/m2), and $\eta_{inverter}$ is the inverter efficiency. The efficiency of the PV array has already been accounted for in the rated capacity as it is based on the actual PV performance at the standard conditions.

The amount of solar irradiance (G_T) received by the solar panels is calculated based on the commonly used isotropic model defined by the following equation [32]

$$G_T = G_b R_b + G_d \left(\frac{1+\cos(\beta)}{2}\right) + G\rho_g \left(\frac{1-\cos(\beta)}{2}\right) \tag{15}$$

where, G_b is the beam component of the solar irradiance, G_d is the diffused component, G is the irradiation received by a horizontal plane, R_b is the ratio between the radiation on a tilted plane and a horizontal plane, β is the tilt angle of the solar panel, and ρ_g is the ground albedo or the reflectance coefficient caused by the ground.

The irradiation received by a horizontal plane (G) can be calculated using the sky clearness index (K_T), which is defined as the ratio between the irradiation measured at the sea level, and the extraterrestrial irradiation measured outside the earth's atmosphere [32]

$$G = K_T G_o \tag{16}$$

G_o is the extraterrestrial irradiation defined by the following equation:

$$G_o = G_{on} \cos(\theta_z) \tag{17}$$

where, G_{on} is the extraterrestrial irradiance on a plane normal to the solar beam and θ_z is the zenith angle.

$$G_{on} = G_{sc}\Big(1.000110 + 0.034221\cos(B) + 0.001280\sin(B) + 0.000719\cos(2B) + 0.000077\sin(2B)\Big) \tag{18}$$

where, G_{sc} is the solar constant (1367 Watts/m2) while B is defined as:

$$B = (n-1)\left(\frac{360}{365}\right) \qquad n = day\,of\,the\,year (1 \leq n \leq 365) \tag{19}$$

G_d is calculated using the Orgill and Hollands correlation [32] and G_b is determined by the equation 20

$$G_b = G - G_d \tag{20}$$

The calculation of the solar irradiance followed by this study can be summarized as follows:

- Predict the extraterrestrial irradiance Go based on the date, time, and site parameters (i.e. latitude, longitude, altitude). Derive the necessary data about solar angles and solar time.

- Derive the clearness index KT based on atmospheric conditions. This value is related to the cloud cover. According to [33], cloud cover is influenced by humidity more than by temperature. Thus, for this study, a regression model was created to derive K_T as a function of humidity.
- Derive clear sky irradiance on horizontal plane (i.e. $G = K_T G_o$)
- Derive the diffused component G_d using Orgill and Hollands correlation
- Calculate the beam component (i.e. $G_b = G - G_d$)
- Calculate R_b (i.e. $R_b = cos(\theta)/cos(\theta_z)$.
- Finally calculate the solar Irradiance G_T

The complete details of calculating the necessary angles and solar time conversion is discussed by [32].

2.3.2. Wind Power Model

The collective performance of several wind turbines depends on the output of each individual turbine which is expressed by its designed power curve. This direct estimate approach has been proven to be more accurate and reliable than other methods such as curve fitting and regression [34].

In this study, the model of an individual wind turbine shall be used to predict the power output of a collection of behind-the-meter wind turbines. Such model specifies the cut-in speed at which the turbine starts to produce electricity, rated speed at which the turbine produces the rated capacity, and cut-out speed (i.e. maximum speed) at which the turbine has to stop its operation for safety purposes. The typical mathematical expression of the wind power curve is [30]:

$$P(v) = P_{capacity} * \begin{pmatrix} 0 & 0 \leq v \leq v_i \\ a + bv^3 & v_i \leq v \leq v_r \\ 1 & v_r \leq v \leq v_o \\ 0 & v > v_r \end{pmatrix} \tag{21}$$

where, v is the wind velocity, v_i is the cut-in speed, v_r is the rated speed, and v_o is the cut-out speed. The coefficients a and b are defined as:

$$a = \frac{v_i^3}{v_i^3 - v_r^3} \qquad b = \frac{1}{v_r^3 - v_i^3} \tag{22}$$

To account the variations wind effectiveness due to air density, the average velocity should be adjusted [34]:

$$v = v_{ave}(\frac{\rho}{\rho_o})^{\frac{1}{3}} \tag{23}$$

where, v_{ave} is the average wind speed, ρ is the current density of air in the site, and ρ_o is the density of air based on the ISO standard atmosphere ($1.225 kg/m^3$).

3. Load Flexibility, Prioritization, and Power Allocation

3.1. Flexibility of Each Type of Flexible Load

The flexible loads are prioritized based on their capacity to accept delays in operation. In this context, the urgent non-thermostatically controlled loads that has direct interaction with the end-user and needs to start immediately after the switch is turned on would have zero flexibility. Therefore, such types of loads should have the highest priority and be served first. This is to enable the end-users to do their daily routines with minimal disruptions from the DR program.

Other load types as discussed above, have their own degree of flexibility which shall be defined and discussed in this section.

3.1.1. Flexibility of TCLs

TCLs are prioritized according to the temperature deviation from the setpoint. Loads with higher values of deviation shall be served first. The common thermostat control of a TCL allows a certain deviation from the setpoint (i.e. deadband). The flexibility of TCLs in cooling and heating mode are illustrated in figures 3 and 4. The flexibility of a TCL in cooling mode can be calculated as:

$$f = \frac{T_{max} - T_a}{T_{max} - (T_s - tolerance)} \tag{24}$$

where, T_s is the setpoint temperature, T_a is the current temperature inside the thermal zone, and T_{max} is the maximum temperature for T_a which can be specified in the contract agreement with the end-users.

Based on the flexibility equation, there is a possibility that the flexibility value would be negative. This happens when the temperature inside the thermal zone has exceeded the contracted maximum temperature limit T_{max}, a scenario that should be avoided. Hence, in the power allocation algorithm, the goal is keep the flexibility values from being negative.

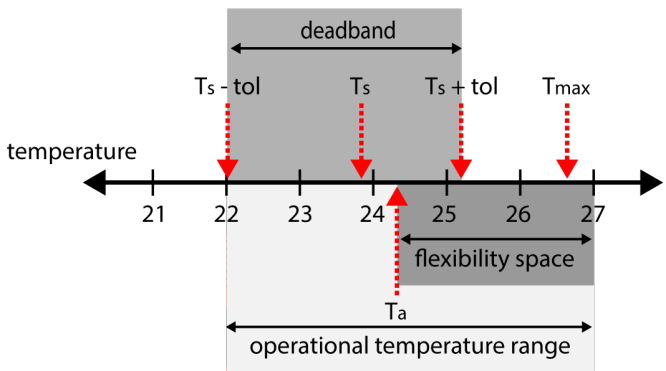


Figure 3: Flexibility of thermostatically controlled loads in cooling mode

Similarly, the flexibility of a TCL in heating mode can be derived as:

$$f = \frac{T_a - T_{min}}{(T_s + tolerance) - T_{min}} \tag{25}$$

where, T_{min} is the minimum temperature contracted with the end-user during the agreement to participate in the DR program. Negative values of flexibility for TCLs in heating mode means that the internal temperature has gone lower than the contracted lower temperature limit.

3.1.2. Flexibility of Battery-based Loads

Battery-based loads like electric vehicles and grid energy storage are prioritized according to the urgency of charging completion. This can be expressed as:

$$f = \frac{(t_{use} - t) - (t_f - t)}{t_{use} - t} \tag{26}$$

where, f is the flexibility, t is the current time, t_f is the expected time to finish charging if charging is started at current time considering its current status, and t_{use} is the time when the unit shall be used and should already be fully charged. This should be specified by the end-user upon plugging in the charger, hence, the charging stations should be able to take this information from the customers.

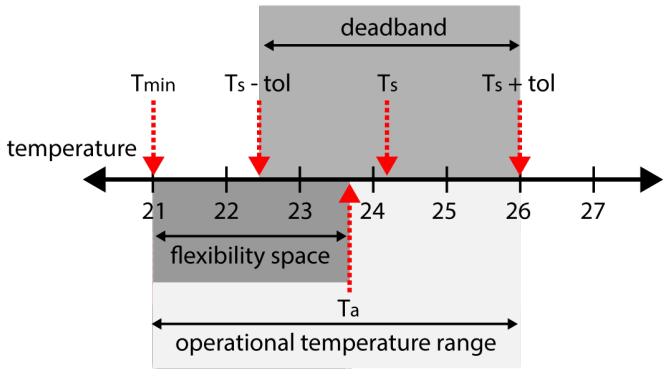


Figure 4: Flexibility of thermostatically controlled loads in heating mode

In this study, battery-based loads are assumed to be interruptible so that whenever there is insufficient power, charging should cease, consequently moving the predicted finish time closer to the latest finish time. Meanwhile, whenever excess power is available, charging should start automatically. Figure 5 illustrates the flexibility of the battery-based loads which is based on delayed charging.

This study does not include fully modulated EV charging or bi-directional vehicle-to-grid (V2G) charging, because the charging control technology necessary to achieve these functionalities is not commonly available at this time. To be reflective of current technology offerings, only binary interruptible charging control is included in the model. The introduction of more sophisticated EV charging controls would of course provide additional DR flexibility from a resource that contained significant EV deployment, though such a scenario is not modeled in this study.

3.1.3. Flexibility of Non-Urgent Non-TCLs

The non-urgent non-TCL loads can be delayed, hence it offers a flexibility space between the predicted finish time based on the duration of its operation and the required latest time to finish.

$$f = \frac{t_L - t_f}{t_L - t} \tag{27}$$

where, t_L is the latest finish time, t_f is the predicted finish time, and t is the current time. Here the predicted finish time can be calculated as:

$$t_f = t + \Delta t_{duration} \tag{28}$$

Similar to the flexibility function of the battery-based loads the flexibility of the non-urgent non-TCLs already take into account the user's requirement for the operation to finish. It is important to note that unlike battery-based loads NNTCLs can not be interrupted once started. The flexibility value may reach a negative value which means that the expected finish time is already beyond the required latest time to finish. Thus, the constraint in this algorithm is to avoid reaching negative values for the flexibility.

3.2. Power Allocation

After calculating the flexibility of the loads, they are prioritized according to their flexibility value. The available power is then allocated to the loads starting from those with lower flexibility. The power allocation is constrained to the limit of the available power which is subject to the DER output and the market price. Hence, any instantaneous fluctuations caused by the DER is automatically absorbed by the aggregated loads

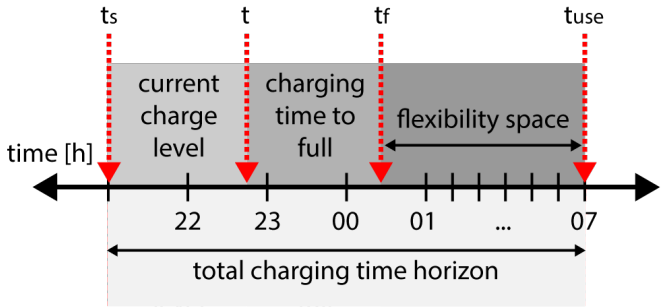


Figure 5: Charging flexibility of battery-based loads

by shedding off the demands with higher flexibility. Such loads could be TCLs that are still far from the temperature limit, battery-based loads which still have sufficient time to be fully charged before they are to be used, or non-urgent loads that can be shifted to operate in off-peak times.

An example of load prioritization and power allocation is presented in Table 1. The loads are sorted according to their flexibility. Power is allocated first to loads with lower flexibility values and they are given the permission to turn on. Notice that load 6 is also allowed to turn on even its flexibility value is higher than load 7 and load 8. This happens because the power demand of load 6 is small enough that it can still be served by the remaining available power that the generators can supply. The algorithm scans the entire list of loads until all of the available power has been allocated without exceeding the limit.

Table 1: Power allocation to flexible loads after prioritization

Load ID	Demand $[Watts]$	Flexibility Value	Permission Status
5	1061.1	0.287	ON
2	807.5	0.334	ON
3	907.3	0.368	ON
1	2340.2	0.472	ON
8	432.7	0.477	ON
9	540.2	0.489	ON
10	659.1	0.632	OFF
7	1354.3	0.682	OFF
6	120.8	0.723	ON
4	2340.8	0.782	OFF

After the status of each load are determined, total electricity cost limit is determined by the product of the price threshold and the sum of all load demands that were permitted to turn on. Such loads are rescanned and their costs (i.e. demand x price) are aggregated with the constraint of not exceeding the total electricity cost limit. Loads that did not make it to the cut-off will be reassigned with an off status, hence it is not permitted to run for the next time increment.

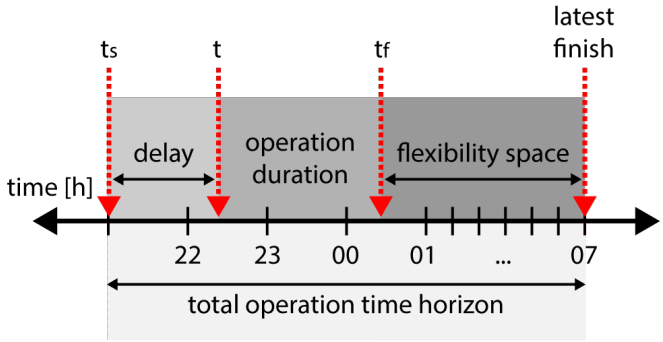


Figure 6: Flexibility of non-urgent non-thermostatically controlled loads

3.3. Simulation

In this section, the viability of the proposed aggregation algorithm is demonstrated via simulation. Figure 7 shows the proposed aggregation algorithm as implemented in Python. This study does not simulate the individual models of loads which were already validated by other authors [29, 32, 34]. The focus of this study is to validate the proposed aggregation methodology. Simulations are conducted on a set of buildings with varying load characteristics.

A hypothetical population of HVACs, refrigerators, freezers, water heaters, EVs, energy storage, and non-TCL loads were generated using a Gaussian distribution of parameters (e.g. area, volume, insulation, U_a, U_m, C_a, C_m, etc.). Then, the time varying parameters were initialized. The parameters are then updated using the corresponding weather conditions of that particular time step. With such parameters the status of the loads were analysed and their flexibilities are calculated. The flexibilities of all the loads are then ranked together during the prioritization process. Loads with lower flexibility are then served first during the power allocation process. The allocated power is maximized to serve as much load requests as possible while following the constraint that it should not exceed the power supply limit. During this process the aggregated demand response is calculated as the total load demand minus the total allocated power. Next, the behaviour of the loads is emulated by inputting the current parameters to their physical models. The resulting values of the variables being monitored (i.e. temperature, charge level, job status) are recorded while the real-time data are updated. The algorithm proceeds to the next time step and repeat the process until it reaches the last time step specified at the beginning of the simulation.

The weather database used for this simulation is the Typical Meteorological Year 3 (TMY3) developed by National Renewable Energy Laboratory. Meanwhile, the typical load demand for a representative building in Los Angeles, California, USA is based on the datasets compiled by openei.org that is associated with the weather database TMY3.

3.3.1. Assumptions for TCLs

Thermal parameter values used for the ETP models of different TCLs were taken from different publications, or assumed [10, 35, 36]. The assumed parameters for HVAC units are shown in Table 2.

The heat gains/losses due to ventilation and leakages in the building are taken into account by assuming that the air change per hour (CPH) is 2.0. To calculate the air mass flow, a regression model for air specific heat capacity (C_p) and air density were created as a function of the outside air temperature.

For refrigerators and freezers, it is assumed that the units are located inside the building, hence, the ambient temperature is the air temperature simulated by the HVAC units. The air changes due to the

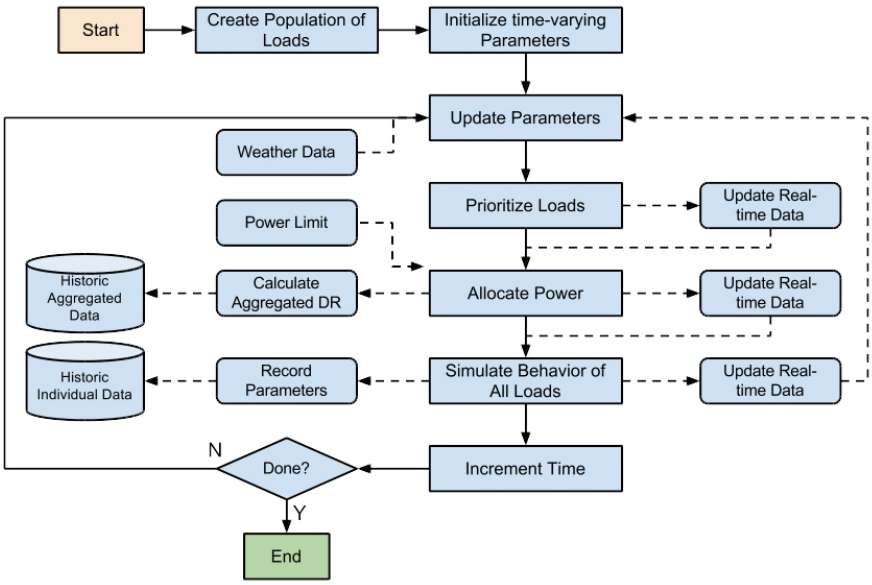


Figure 7: Proposed algorithm for aggregation of flexible loads

Table 2: Range of values for the parameters of the HVAC models

Parameter	Lower	Upper	Average
Area [m2]	139	325	186
U_a $[W/^oC]$	0.85*Area	1.98*Area	211
U_m $[W/^oC]$	0.85*Area	1.98*Area	211
C_a $[J/^oC]$	3025*Area	3025*Area	3025*Area
C_m $[J/^oC]$	900*Area	900*Area	900*Area
$T_{setpoint}$ $[^oC]$	23.9	24.4	23.9
Tolerance $[^oC]$	0.28	1.1	0.56
T_{max} $[^oC]$	26	28	27
T_{min} $[^oC]$	15	17	16
Heating $[W]$	56.75*Area	56.75*Area	56.75*Area
Cooling $[W]$	56.78*Area	113.56*Area	14064

opening of the refrigerator and freezer door are accounted for in a constant value of air mass flow (i.e. $0.000001kg/s$). Since the ambient temperature as well as the temperature inside the units are acceptably stable, the specific heat capacity C_P is assumed constant (i.e., $1005J/kg^oC$). Meanwhile, the coefficient of heat transfer and the heat capacity of air and food materials are assumed to be the same for both refrigerators and freezers.

Though there are some entries in Tables 3, 4, and 5 with only one value for Lower, Upper, and Average values, they are commonly multiplied by a certain factor e.g. volume, thus, the actual values still vary according to the Gaussian distribution. Tolerance values set the limit above or below the temperature setpoint. In that case, the operating deadband of TCLs in this study is from $(T_{setpoint} - Tolerance)$ to $(T_{setpoint} + Tolerance)$. For water heaters, the specific heat capacity of water is assumed to be constant (i.e., $4186J/kg^oC$). The effect of the internal mass is neglected in the ETP model for water heaters.

The mass flow variations caused by usage of hot water will have an impact on the performance of water heaters. In this study, a regression model for water usage has been created based on the water heater demand from the database for typical residential loads. The regression model is a function of the time of the day, i.e. 0 to 24. To attain a better accuracy (i.e. for per minute values), the total time horizon is divided

Table 3: Range of values for the parameters of the refrigerator models

Parameter	Lower	Upper	Average
Volume $[m^3]$	0.324	0.72	0.5
U_a $[W/^oC]$	2.27 * Volume		
U_m $[W/^oC]$	66.67* Volume		
C_a $[J/^oC]$	1210 * Volume * Air density		
C_m $[J/^oC]$	1500 * Volume		
$T_{setpoint}$ $[^oC]$	1.7	4.0	2.8
Tolerance $[^oC]$	0.28	1.1	0.56
T_{max} $[^oC]$	8	15	12
Cooling [W]	350 * Volume		

Table 4: Range of values for the parameters of the freezer models

Parameter	Lower	Upper	Average
Volume $[m^3]$	0.24	0.72	0.3
U_a $[W/^oC]$	2.27 * Volume		
U_m $[W/^oC]$	66.67 * Volume		
C_a $[J/^oC]$	1210 * Volume * Air density		
C_m $[J/^oC]$	1500 * Volume		
$T_{setpoint}$ $[^oC]$	23.0	18.0	21.0
Tolerance $[^oC]$	0.28	1.1	0.56
T_{max} $[^oC]$	0	4	2
Cooling [W]	350 * Volume		
C_p $[J/kg^oC]$	1005		

into six parts and a regression model is created for each part.

$$F_m = 0.0003 * \begin{pmatrix} 0.019 & 0 \leq t \leq 4 \\ 0.6(t-4)+0.25 & 4 < t \leq 8 \\ -0.1(t-8)+0.99 & 8 < t \leq 12 \\ -0.01(t-12)+0.44 & 12 < t \leq 16 \\ 0.06(t-16)+0.14 & 16 < t \leq 20 \\ -0.14(t-20)+0.8 & 20 < t \leq 24 \end{pmatrix} \tag{29}$$

where, F_m is the mass flow, and t is the time in hours from 0 to 24 (see Figure 8). Random time delay (i.e. + 30 minutes) is introduced on each unit to mimic the realistic variation of load schedule.

3.3.2. Assumptions for Battery-based loads

The first assumption made for battery-based loads is that their charging process can be interrupted. Other assumptions for the different parameters of the battery model such as the capacity, and charging rate are based on actual values of existing EVs in the market. The types or brands of units which are commonly bought by customers are also considered as they depict the realistic average battery capacity and charging rate of EVs that connects to the grid. The following range of parameter values were derived for electric vehicles, as presented in Table 6.

The status of EV batteries are also simulated based on the typical driving schedule of the users, i.e. they leave the house at around 05:00 to 07:00 and return home at around 17:00 to 21:00. It is assumed that at the start of charging (i.e. when the EV is plugged in), the users would specify the required latest time to finish charging. On the other hand the battery-based energy storage parameters are shown in Table 7. The discharging rate is assumed to be equal to the charging rate.

Table 5: Range of values for the parameters of the water heater models

Parameter	Lower	Upper	Average
Volume $[m^3]$	0.0075	0.03	0.0225
U_w $[W/^oC]$	75 * Volume		
C_w $[J/^oC]$	4186 * Volume * Water density		
$T_{setpoint}$ $[^oC]$	50.0	65.0	60.0
Tolerance $[^oC]$	0.28	1.1	0.56
T_{min} $[^oC]$	35	45	40
Heating [W]	200000 * Volume		
Flow $[kg/s]$	0.0003 * Water Usage		
C_p $[J/kg^oC]$	4186		

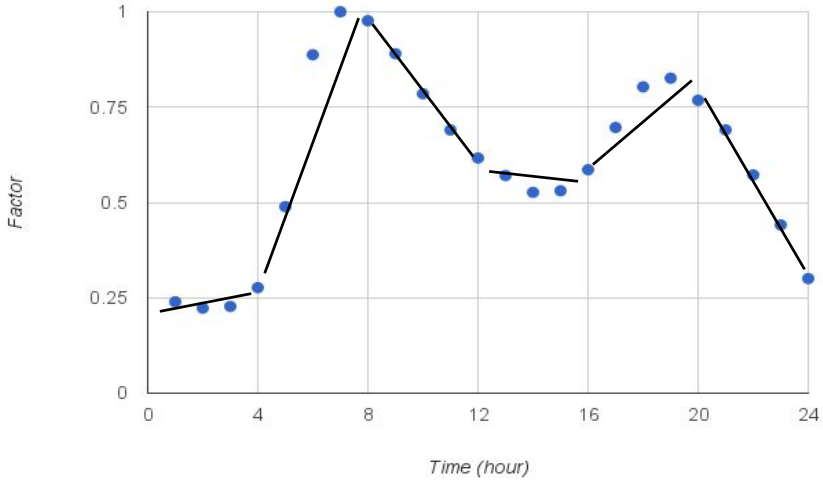


Figure 8: Graphical representation of regression functions for hot water usage

Energy storage devices are not bounded by a particular schedule for charging start and the latest time to finish charging. Instead, they operate according to the balance between the power supply and demand. If demand exceeds supply, the batteries will discharge to help in supplying power. Conversely, when demand is lesser than supply, the battery will switch to charging mode. It is important to note that in discharging mode, the state of charge should stay above 20%. Below this limit, the discharging should stop to minimize battery degradation [37]. Flexibility values are fixed at 9999 just to emphasize that in load demand perspective, they are of least priority. In addition, it is assumed that 50% of the buildings have battery-based energy storage device.

3.3.3. Assumptions for Urgent Non-TCL Loads

Urgent non-TCL loads considered in this study include the lighting and miscellaneous loads (i.e. appliances for cooking, cleaning, entertainment,etc.). Regression models were created with similar approached used to model the hot water usage. The total time horizon is divided into four parts and regression models for each part were developed. The typical hourly operating schedules are presented in Figure 9 to predict the power demand of lighting loads and appliance loads.

where, P_{light} is the power demand by lighting loads $[kW]$, P_{misc} is the power demand by appliance loads $[kW]$, and t is the time of the day $[h]$. The total urgent load demand is the sum of P_{light} and P_{misc}. These loads are also assumed to be converted into heat which contribute to the internal heat gain that should be included in the HVAC model.

Table 6: Range of values for the parameters for electric vehicles

Parameter	Lower	Upper	Average
Capacity $[kWh]$	5.2	85	24
Charging Rate $[kW]$	3.3	20	3.3
Charging Efficiency [%]	0.75	0.9	0.8
Charging Start $[h]$	17	21	19.5
Latest Finish $[h]$	5	7	6

Table 7: Range of values for the parameters of energy storage

Parameter	Lower	Upper	Average
Capacity $[kWh]$	7	90	10
Charging Rate $[kW]$	2.0	3.3	2.0
Charging Efficiency [%]	0.85	0.95	0.92

With regarding to the simulation sequence, the following steps of action were performed to model difference grid scenarios.

1. Calculation for the existing grid power capacity. The power capacity of the existing grid was calculated by simulating a set of hypothetical loads without electric vehicles and energy storage devices. This is to check the viability of the algorithm to provide DR and delay the necessary infrastructure upgrades and construction of additional power plants while EVs and energy storage devices are integrated into the existing grid. The system capacity is determined by the following equation. The typical diversity factor (i.e. 3) for residential sector was adopted in this simulation.

$$System\,Capacity = \frac{\sum(Individual\,Peak\,Loads)}{Diversity Factor} \tag{30}$$

2. Getting the baseline load demand. In this simulation, EVs and energy storage devices were added to the total load curve. Two baseline load curves were established: one curve includes EVs but excludes the energy storage devices; the second curve includes both the EVs and battery-based energy storage devices. This is to check if energy storage devices are necessary or the EVs are already sufficient to provide more flexibility to the grid.
3. Load shedding simulation: The grid capacity which was derived in simulation 1 was used as the power limit for the baseline curve in simulation 2. The maximum increase or decrease of power supply at each time step (i.e. 1 minute) was determined by adding the maximum ramping rate of the existing generators and the DER forecast at such period. It was assumed that the traditional generators have a collective ramping rate of 1% of its capacity per minute [38]. The upper limit of the entire simulation period (i.e. 24 hours) is the maximum capacity of the traditional generators plus the maximum DER output at a particular time. The total load demand should not exceed the power limit, hence, some loads are automatically shifted by the control algorithm during the power limiting hours.
4. Response to DER fluctuations. A sudden drop of DER output was simulated by reducing the DER electricity production by 80%. The algorithm automatically reduces the load demand to maintain the power balance. It was assumed that 50% of the buildings has installed a rooftop solar PV system, and the roof area used for solar PV varies from 25% to 70% of building's floor area. On the other hand, the installed wind capacity was assumed to be 10% of the grid capacity.
5. Response to market fluctuations. A hypothetical variation of market price was simulated and the threshold price was chosen to be \$0.15 per kWh. The demand automatically reduces when the price exceeds the threshold price level.
6. Impact of the energy storage devices. It was assumed that each building has an energy storage device that can help in managing the grid. State of charge of the batteries were not allowed to discharge below 20% of its capacity. This is to consider the safety operation of the batteries and minimize degradation.

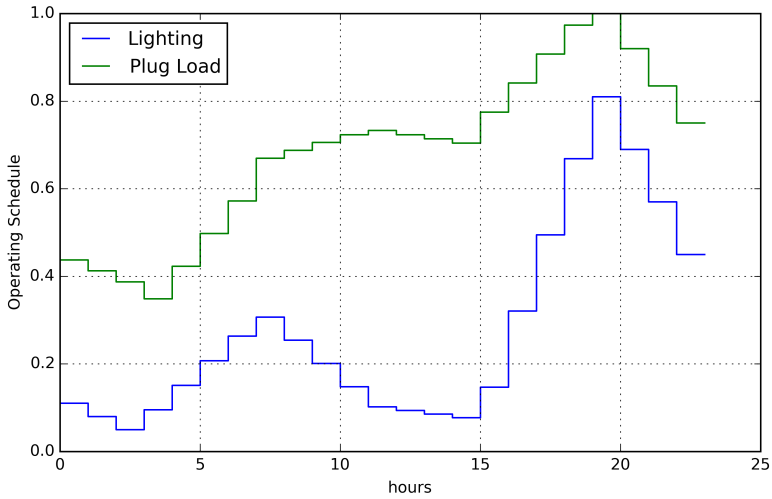


Figure 9: Hourly Operating Schedules of Urgent Non-TCL Loads

4. Results

For the simulation results presented in this section battery-based energy storage technologies are not considered. Additionally, the electric vehicles act solely as a load and do not contribute in supplying power to the grid. This is done to emphasize the ability of the aggregation methodology in maintaining the power balance without relying on dedicated, dispatchable resources like energy storage devices.

To provide additional context for how the controller will utilize and influence the aggregate demand Figure 10 shows the original load demand, i.e., before the proposed algorithm was applied. It also shows the aggregated load demand as typically seen from the substation or aggregation point. Each demand load-type is summed and grouped by color. As expected, the dominant loads are cooling (i.e. blue) during daytime hours and electric vehicle charging (i.e. green) overnight. The high cooling demand is logical since the outside air temperature (OAT) is high. No other load exhibits and obvious diurnal pattern. As shown in Figure 10, market energy prices are presented to quantify each type of flexible load' response to higher energy price. Market prices will be relevant to scenarios presented in subsequent sections.

4.1. Scenario 1: System Peak Mitigation

The simulation results presented in Figure 11 illustrate the response of the algorithm to mitigate large demand spike which occurs between approximately 18:00 and 23:00 in the base case scenario. To accomplish this, the controller relies predominantly on shifting EV charging, while also utilizing flexibility in refrigeration loads and NNTCLS. The magnitude of each type of flexible load contribution is determined by the load prioritization algorithm. The power allocation algorithm prevents the load demand to exceed the available power supply. It can also be observed that there is no rebound of power demand after the peak hours for TCL loads, which is commonly seen after a DR event in other DR programs (e.g. Time-of-Use).

With such demand response, the performance of each load is slightly affected by participating in DR, but does not exceed the limits of its specified operation conditions, particularly temperature for TCLs and completion status for non-TCLs. Figures 12 shows the operating temperatures of HVACs, refrigerators, water heaters, and freezers during the DR event. As shown in Figure 12 the TCL temperatures are within the limits of the normal operation. The charge of the electric vehicles and the job status of the NNTCLs are also done before the required time as shown in 13.

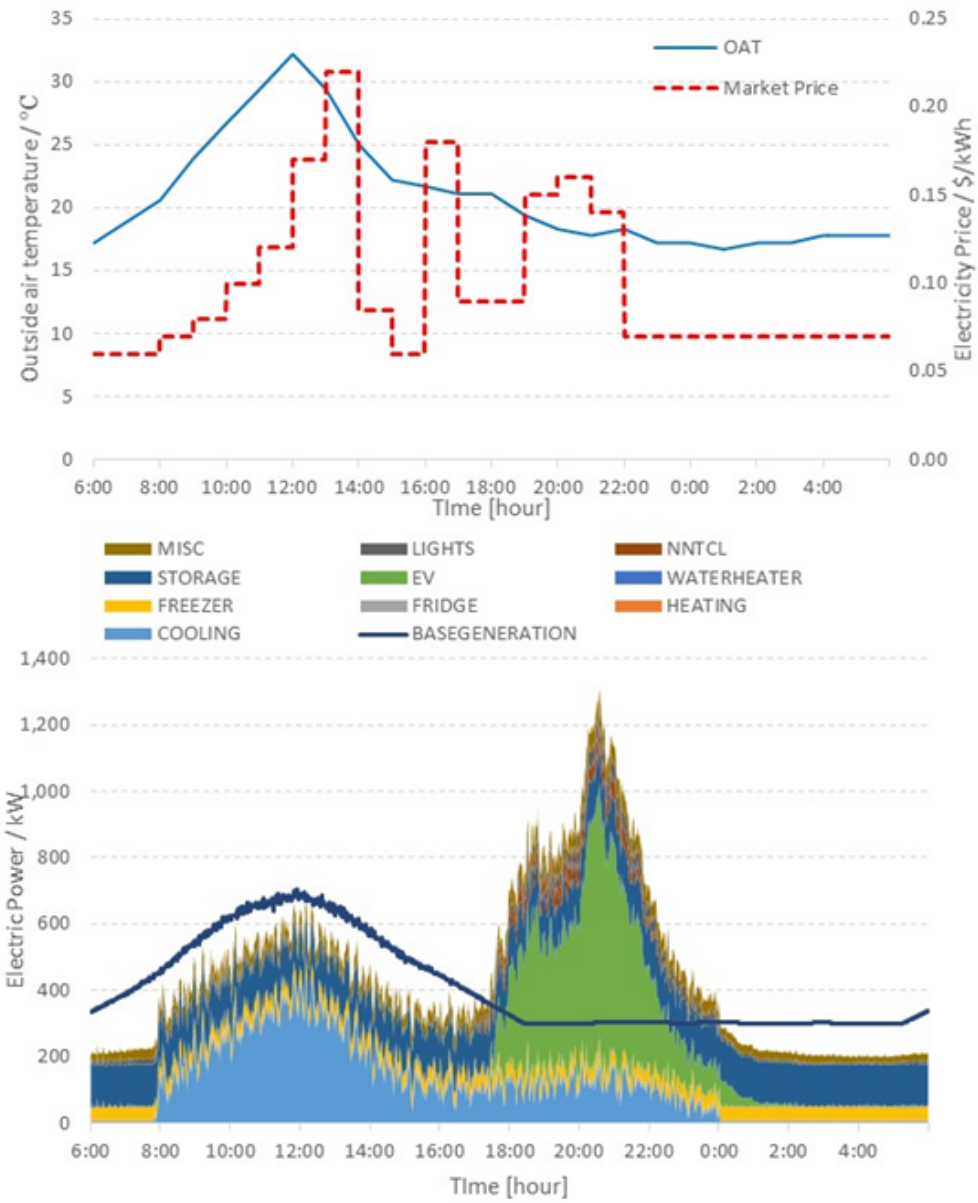


Figure 10: Flexible loads' aggregated demand and renewable generation on a summer day

4.2. Scenario 2: DER Disruptions

Figure 14 illustrates how the controller is able to adjust the total load demand instantaneously in response to DER fluctuations. In this figure, PV output within the system decreased by 80% for 1 hour (i.e. from 11:30 to 12:30). During such an event, traditional generators will ramp up to meet the newly increased net demand, however it is not capable of providing sufficient power instantaneously. Under these conditions, the ability of the controller to provide fast DR is highly valuable. Figure 14 demonstates how the controller utilized each load-type to respond the loss of renewable generations. In this scenario, the controller is also reducing the system peaks caused by EV charging during evening hours. It can be seen that HVACs, freezers, and refrigerators provide all the load shed capacity.

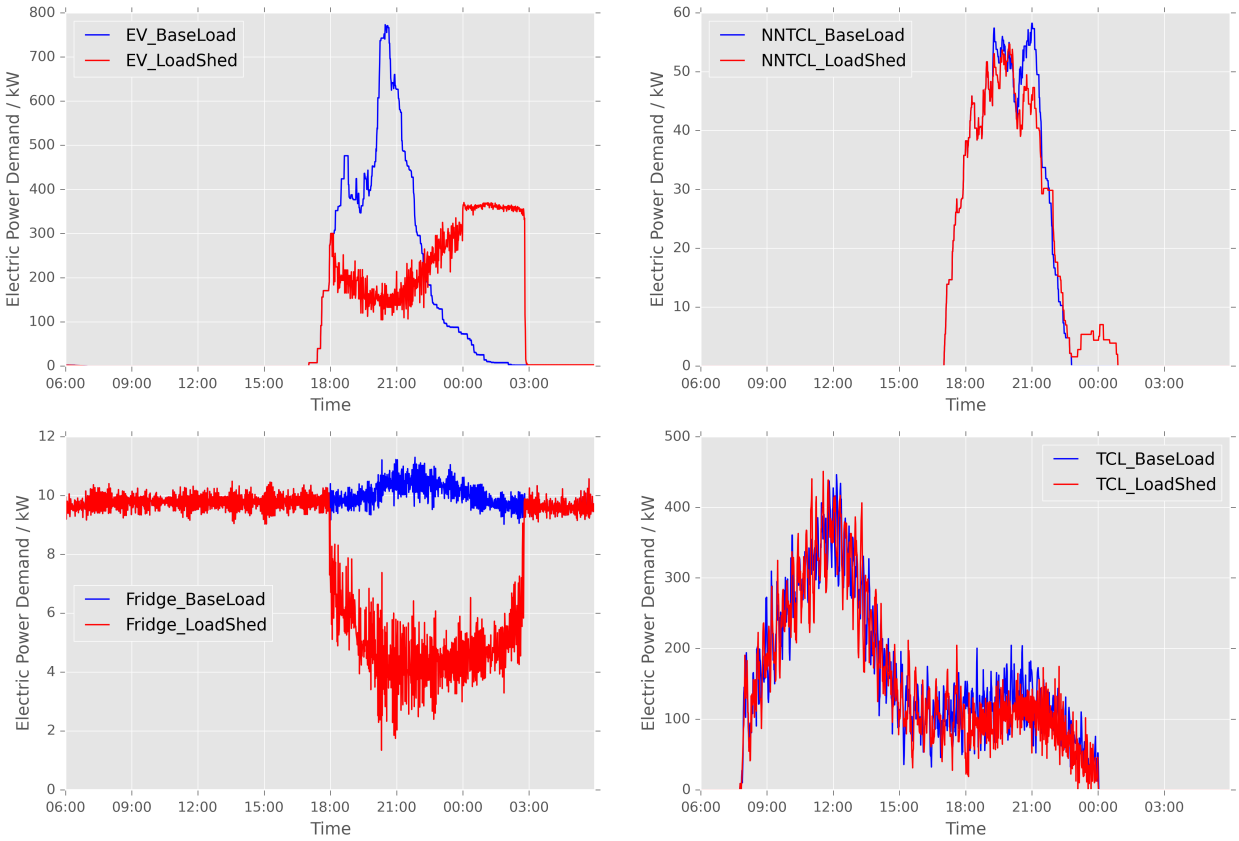


Figure 11: Load shedding during peak hours

4.3. Scenario 3: Market Price Fluctuations

Finally, this scenario introduces fluctations in the market price for energy, as illustrated in Figure 10. The results from the simulated controller for this scenario are given in Figure 15. A threshold price has been set at 0.15 $\$/kWh$. When the market price exceeds the threshold price, the controller responds by reducing consumption similarly as when under a DR event, while allowing consumption to increase when market prices remain below the threshold. The high-price periods (i.e. from 13:00 to 15:00 and from 17:00 to 18:00) are highlighted in the figure. As with scenarios 1 and 2, the overall demand is still subject to the limitation of power supply. Meanwhile, the load shed capacities from EVs, NNTCL, fridges, and freezers are deployed to prevent the aggregated load demand to exceed the supply limit. The operating conditions of the loads are slightly affected but within prescribed limits.

5. Discussion, Conclusion and Future Work

This paper provides the details of the development of an algorithm for aggregation of flexible loads for demand response application at the substation level. Its viability has been demonstrated by simulations of the different flexible load models aggregated by the proposed algorithm. Results show that the algorithm can effectively mitigate system demand peaks with minimal compromise in the load operations. It is also observed that sudden spikes on the load demand (rebound effect) commonly observed in other DR programs such as Time-Of-Use are avoided in this methodology. The premise is that the algorithm should run in the background during normal operations in the same manner as spinning reserves operate. The algorithm can react instantaneously and can provide fast demand response or auto-DR to absorb grid disruptions

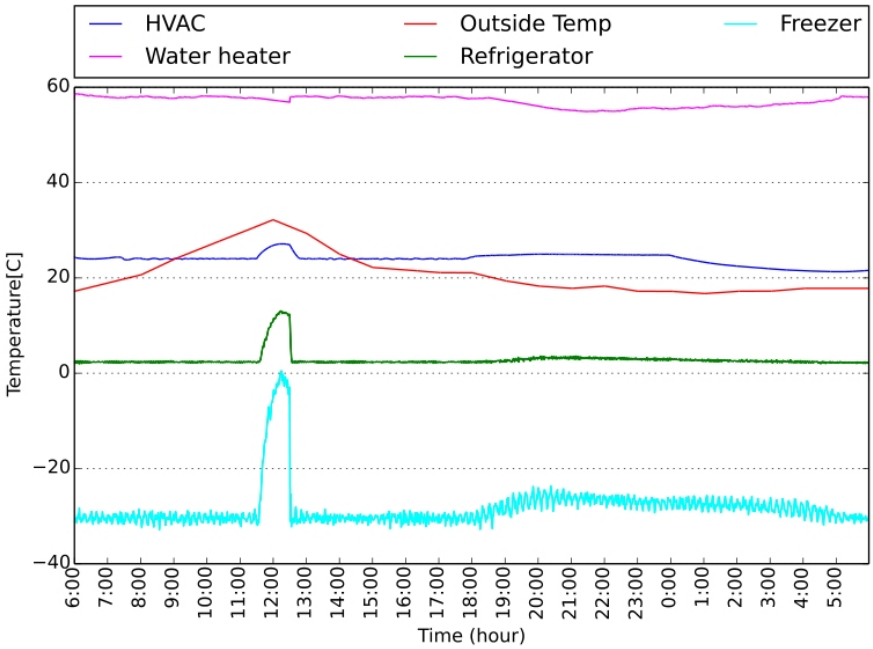


Figure 12: Status of flexible loads to DER output disruption and load shedding

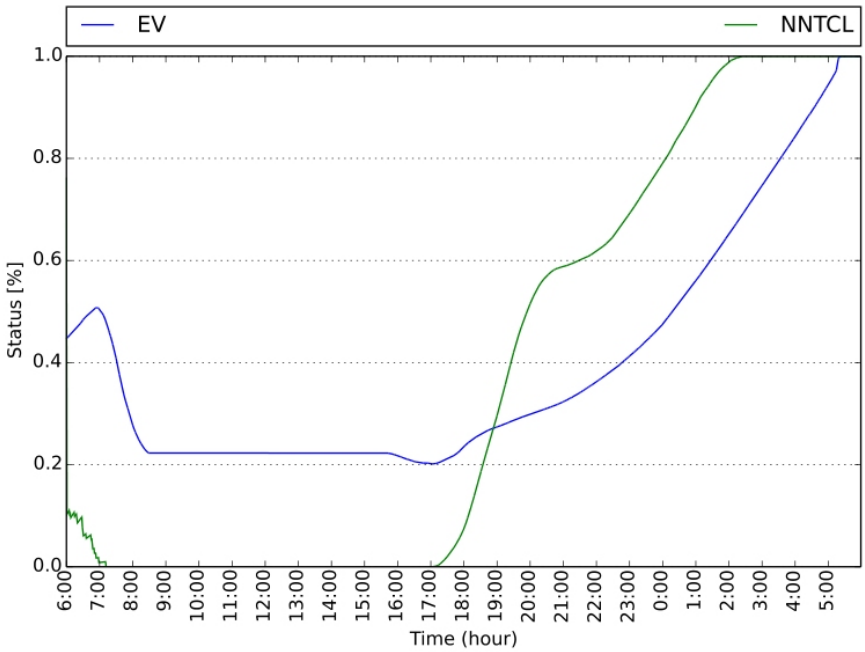


Figure 13: EV and NNTCLs' status in response to system peaks and power limitations

caused by fluctuations in DER output or market price of electricity. With all these potential advantages, the proposed methodology proves to be a good candidate to assume the role of ancillary services and assist in accelerating the adoption of the intermittent renewable energy resources.

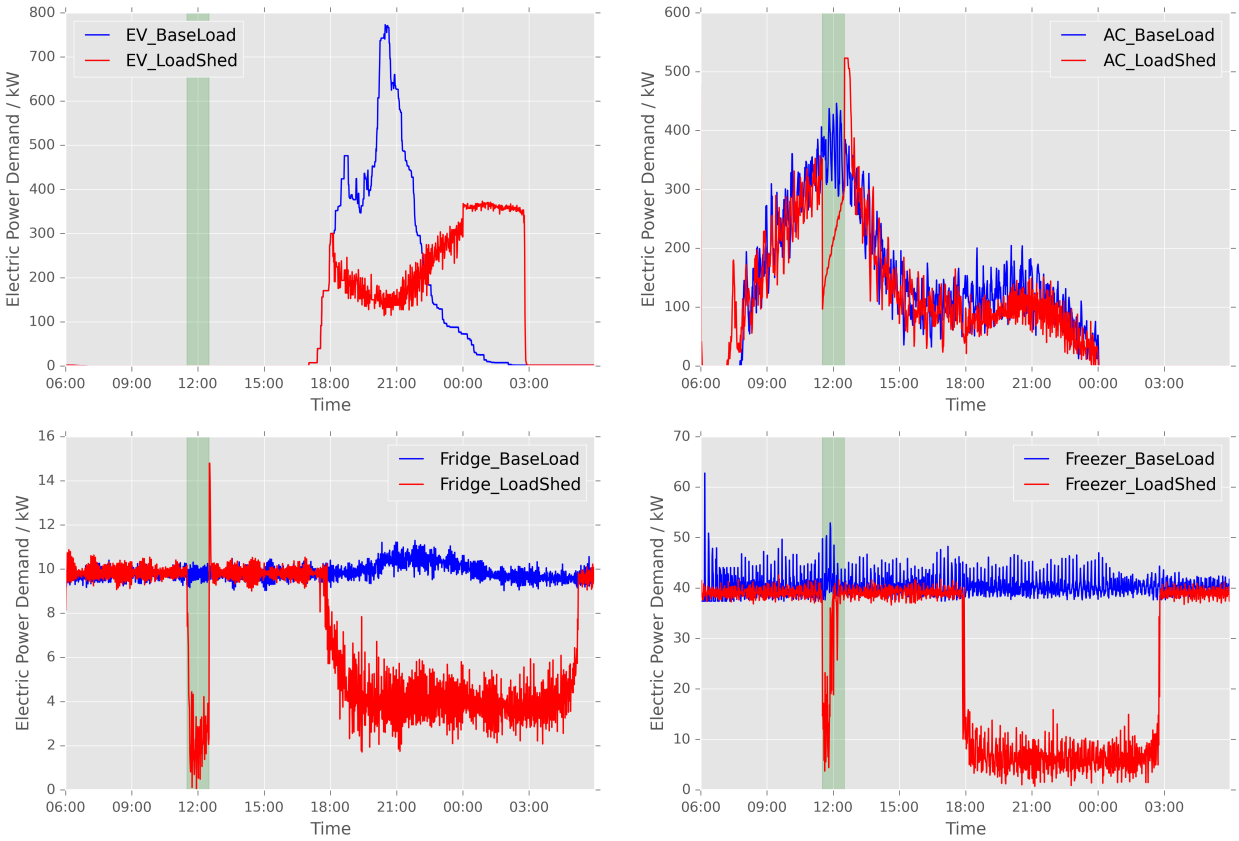


Figure 14: Response of flexible loads to DER output disruption and load shedding

In comparison with previous work, we create a suite of bottom-up physical models of each energy system component for modeling the DR potential from residential sector. Specifically, those models are grounded in the physical constraints for thermostatically controlled loads (i.e. HVAC, refrigerators, electric water heaters.), EV and storage, and distributed PV and wind generations. In addition, the proposed power allocation algorithm for providing demand response resources is unique base on each energy system component's load flexibility. Lastly, it is expected to provide an accurate prediction of DR potential for day-ahead planning simulations in a large scale due to the sufficiently low computational intensity. With regarding to the limitation of this framework, we need to know a lot about the composition of the feeder system to build up the model. In addition, we also need to collect a lot of the equipment stock and conditions to build model for each system component under the feeder system, because those models vary a lot by a number of parameters, such as location, weather, building stocks, EV driving and storage usage behaviors. In addition to those uncertainties in model buildups, the computational intensity may be too high to generate for higher granularity or real-time simulation in a large scale.

Further studies of this methodology may include the improvement of the models of different flexible loads and optimization algorithm to close the gap between the supply and demand without the help of energy storage devices. In actual implementation of this technology, models of the flexible loads may no longer be necessary. Instead of simulating the models, the loads may directly send its current status to the central controller periodically. It is also recommended to implement the same algorithm in grid simulation software to verify its viability when electrical behaviours of the grid is taken into account.

Development and analysis of business models based on this type of aggregation model and utility-customer interactions would also be of interest. This might include investigations of the necessary market policies, technological requirements, social promotions, and financial aspects needed to support and operate

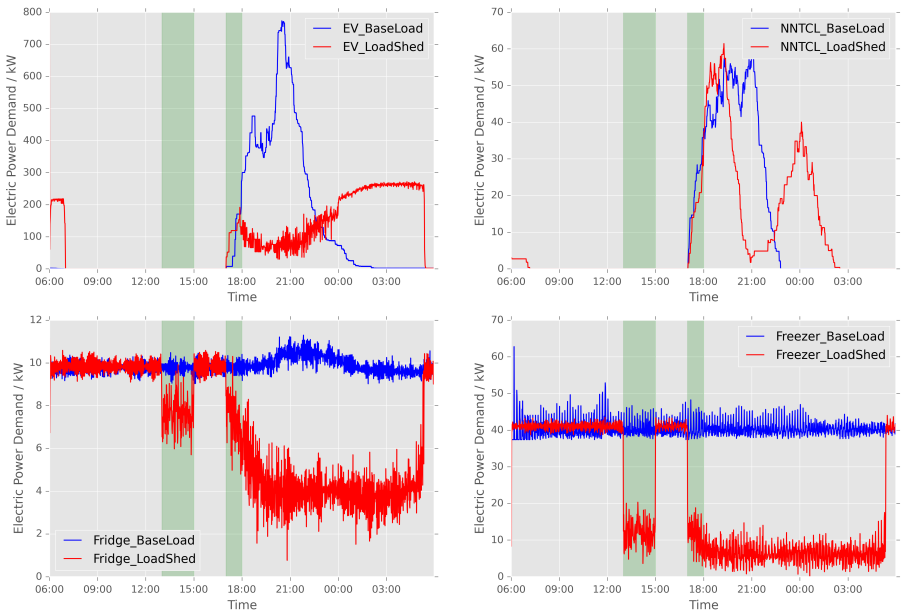


Figure 15: Adjustment of the load demand based on the electricity price

such a model.

6. Acknowledgements

The work described in this study was coordinated by the Grid Integration Group of Lawrence Berkeley National Laboratory and was supported by the State Grid Corporation of China Project (DZN17201300197, Study on Key Technologies for Power and Frequency Control of System with 'Source-Grid-Load' Interactions).

7. References

[1] Ahmad Faruqui, Ryan Hledik, S. S. George, J. Bode, P. Mangasarian, I. Rohmund, G. Wikler, D. Ghosh, and S. Yoshida, "A National Assessment of Demand Response Potential," tech. rep., Federal Energy Regulatory Commission, 2009.

[2] S. Kiliccote, M. a. Piette, J. Mathieu, and K. Parrish, "Findings from Seven Years of Field Performance Data for Automated Demand Response in Commercial Buildings," in *2010 ACEEE Summer Study on Energy Efficiency in Buildings*, 2010.

[3] M. A. Piette, D. Watson, N. Motegi, S. Kiliccote, P. Xu, and D. Watson, "Automated Critical Peak Pricing Field Tests : Program Description and Results," Tech. Rep. August, Lawrence Berkeley National Laboratory, Berkeley, CA, 2007.

[4] P. Cappers, J. MacDonald, C. Goldman, and O. Ma, "An assessment of market and policy barriers for demand response providing ancillary services in U.S. electricity markets," *Energy Policy*, vol. 62, no. March, pp. 1031–1039, 2013.

[5] O. Ma, N. Alkadi, P. Cappers, P. Denholm, J. Dudley, S. Goli, M. Hummon, S. Kiliccote, J. MacDonald, N. Matson, D. Olsen, C. Rose, M. D. Sohn, M. Starke, B. Kirby, and M. O'Malley, "Demand Response for Ancillary Services," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 1988–1995, 2013.

[6] G. Heffner, C. Goldman, and B. Kirby, "Loads Providing Ancillary Services : Review of International Experience," Tech. Rep. May, Lawrence Berkeley National Laboratory, Berkeley, CA, 2007.

[7] D. J. Olsen, N. Matson, M. D. Sohn, C. Rose, J. Dudley, S. Goli, S. Kiliccote, M. Hummon, D. Palchak, J. Jorgeson, P. Denholm, and S. Ma, "Grid Integration of Aggregated Demand Response , Part 1 : Load Availability Profiles and Constraints for the Western Interconnection," Tech. Rep. September, Lawrence Berkeley National Laboratory, Berkeley, CA, 2013.

[8] B. J. Kirby, "Frequency Regulation Basics and Trends," Tech. Rep. December, Oak Ridge National Laboratory, 2004.

[9] David S. Watson, Mary Ann Piette, Osman Sezgen, and Naoya Motegi, "Machine to Machine (M2M) Technology in Demand Responsive Commercial Buildings," in *2004 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA*, 2004.

[10] S. Katipamula and N. Lu, "Evaluation of residential HVAC control strategies for demand response programs," *ASHRAE Transactions*, vol. 112 PART 1, no. June 2015, pp. 535–546, 2006.

[11] S. David, M. Ann, D. S. Watson, S. Kiliccote, N. Motegi, and M. A. Piette, "Strategies for Demand Response in Commercial Buildings," in *2006 ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 287–299, 2006.
[12] M. A. Piette, D. S. Watson, N. Motegi, and N. Bourassa, "Findings from the 2004 fully automated demand response tests in large facilities," Tech. Rep. October, Lawrence Berkeley National Laboratory, Berkeley, CA, 2005.
[13] R. Yin, X. Peng, K. Sila, and M. A. Piette, "Study on Auto-DR and pre-cooling of commercial buildings with thermal mass in California," *Energy and Buildings*, pp. 965–975, 2010.
[14] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi, "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4655–4664, 2013.
[15] P. Denholm, J. Jorgenson, T. Jenkin, D. Palchak, B. Kirby, and M. O. Malley, "The Value of Energy Storage for Grid Applications," tech. rep., National Renewable Energy Laboratory, USA, May 2013.
[16] J. Lopes, F. J. Soares, and P. Almeida, "Integration of Electric Vehicles in the Electric Power System," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 168–183, 2011.
[17] S. Han, S. Han, and K. Sezaki, "Estimation of Achievable Power Capacity From Plug-in Electric Vehicles for V2G Frequency Regulation: Case Studies for Market Participation," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 632–641, 2011.
[18] E. C. Kara, J. S. Macdonald, D. Black, M. Bérges, G. Hug, and S. Kiliccote, "Estimating the benefits of electric vehicle smart charging at non-residential locations: A data-driven approach," *Applied Energy*, vol. 155, pp. 515 – 525, 2015.
[19] H. Lund and W. Kempton, "Integration of renewable energy into the transport and electricity sectors through V2G," *Energy Policy*, vol. 36, no. 9, pp. 3578–3587, 2008.
[20] D. Dallinger, S. Gerda, and M. Wietschel, "Integration of intermittent renewable power supply using grid-connected vehicles – A 2030 case study for California and Germany," *Applied Energy*, vol. 104, pp. 666–682, 2013.
[21] C. Goebel and D. S. Callaway, "Using ICT-Controlled Plug-in Electric Vehicles to Supply Grid Regulation in California at Different Renewable Integration Levels," *IEEE Transactions on Smart Grid*, vol. 4, pp. 729–740, jun 2013.
[22] J. L. Mathieu, M. Dyson, and D. S. Callaway, "Using Residential Electric Loads for Fast Demand Response : The Potential Resource and Revenues , the Costs , and Policy Recommendations," in *Proceedings of the ACEEE Summer Study on Buildings*, pp. 189–203, 2012.
[23] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway, "Energy Arbitrage with Thermostatically Controlled Loads," *European Control Conference*, pp. 2519–2526, 2013.
[24] J. L. Mathieu, M. E. Dyson, and D. S. Callaway, "Resource and revenue potential of California residential load participation in ancillary services," *Energy Policy*, vol. 80, pp. 76–87, 2015.
[25] H. Hao, Y. Lin, A. S. Kowli, P. Barooah, and S. Meyn, "Ancillary Service to the grid through control of fans in commercial Building HVAC systems," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 2066–2074, 2014.
[26] E. C. Kara, M. D. Tabone, J. S. MacDonald, D. S. Callaway, and S. Kiliccote, "Quantifying flexibility of residential thermostatically controlled loads for demand response: a data-driven approach," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pp. 140–147, 2014.
[27] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Aggregate Flexibility of Thermostatically Controlled Loads," *Ieee Transactions on Power Systems, Vol. 30, No. 1*, vol. 30, no. 1, pp. 189–198, 2015.
[28] E. C. Kara, M. Berges, and G. Hug, "Impact of disturbances on modeling of thermostatically controlled loads for demand response," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2560–2568, 2015.
[29] Z. T. Taylor, K. Gowri, and S. Katipamula, "GridLAB-D Technical Support Document: Residential End-Use Module Version 1.0," tech. rep., Pacific Northwest National Laboratory, USA, 2008.
[30] A. Ghasemi, S. S. Mortazavi, and E. Mashhour, "Hourly demand response and battery energy storage for imbalance reduction of smart distribution company embedded with electric vehicles and wind farms," *Renewable Energy*, vol. 85, pp. 124–136, 2016.
[31] X. Geng and P. P. Khargonekar, "Electric vehicles as flexible loads: Algorithms to optimize aggregate behavior," in *2012 IEEE 3rd International Conference on Smart Grid Communications, SmartGridComm 2012*, pp. 430–435, 2012.
[32] J. Duffie and W. Beckman, *Solar Engineering of Thermal Processes: Fourth Edition*, vol. 42. John Wiley and Sons, 2013.
[33] A. M. Tompkins, "Impact of temperature and humidity variability on cloud cover assessed using aircraft data," *Q.J.R. Meteorol. Soc.*, vol. 129, pp. 2151–2170, 2003.
[34] Y.-H. Wan, E. Ela, and K. Orwig, "Development of an Equivalent Wind Plant Power-Curve," in *Proceedings of the WindPower 2010, Dallas, Texas May 23-26*, 2010.
[35] R. Pratt, C. Conner, B. Cooke, and E. Richman, "Metered end-use consumption and load shapes from the ELCAP residential sample of existing homes in the Pacific Northwest," *Energy and Buildings*, vol. 19, no. 3, pp. 179–193, 1993.
[36] W. Sandusky, E. Pearson, N. Miller, R. Crowder, G. Parker, R. Mazzucchi, G. Stokes, J. Thomas, R. Pratt, G. Schuster, M. Halverson, J. Stoops, F. Peterson, R. Gillman, R. Stokes, and S. Hauser, "ELCAP operational experience," *Energy and Buildings*, vol. 19, no. 3, pp. 167–178, 1993.
[37] S. Habib, M. Kamran, and U. Rashid, "Impact analysis of vehicle-to-grid technology and charging strategies of electric vehicles on distribution networks – A review," *Journal of Power Sources*, vol. 277, pp. 205–214, 2015.
[38] B. Kirby and M. Milligan, "A Method and Case Study for Estimating the Ramping Capability of a Control Area or Balancing Authority and Implications for Moderate or High Wind Penetration," *Proceedings of the WindPower 2005*, pp. 1–16, 2005.