Tale of Two Communities, Two Streams of Data and The Duck Curve: Field Testing Results for a Residential Virtual Power Plant Pilot

Michael Zeifman, Fraunhofer USA Center for Manufacturing Innovation
Kurt Roth, Fraunhofer USA Center for Manufacturing Innovation
Hannah Fontenot, Fraunhofer USA Center for Manufacturing Innovation
Donna Leclerc, Sense Labs, Inc.
Brandon Enrique Calderon Granados, Sense Labs, Inc.

ABSTRACT

Our paper presents the results of testing and refining our predictive models and optimization algorithms using real-world data from a load shaping pilot for a cohort of single-family homes. We recruited about fifty homes in two communities, one located in the Northeast and the other in the Pacific Northwest. All participating homes had a central AC unit, about 30% had at least one electric vehicle (EV), and about 30% had a solar PV system. Each home was equipped with an electric sensor that provided disaggregated power time series for major appliances in near-real time and communicating thermostats that provided HVAC runtime information and the indoor environment (temperature and relative humidity). To perform load shaping, we developed two sets of algorithms. The first set of algorithms uses physics-based and/or data-driven self-calibrating models to forecast HVAC loads, PV generation, and EV charging loads, while the second optimizes temperature setpoint schedules and EV charging times to achieve different grid objectives. Both algorithmic sets are computationally efficient and scalable, with the optimization runtime scaling approximately linearly with the number of controlled devices. Whereas our predictive AC and PV models performed well, we found that many EVs were not connected during field events, had a high state of charge when connected, and a sizeable fraction of EVs disconnected during events, making EVs an unreliable resource in our experiments. Furthermore, the naïve optimization approach used in the pilot was not able to reliably find optimal dispatch solutions. We conducted a virtual pilot using semi-synthetic data sampled from the pilot homes. It clearly demonstrated the value of the computationally efficient optimization algorithm for identifying optimal/near-optimal dispatch schedules for a community of homes. as it effectively flattened the aggregate load profile over a 12-hour time frame, while simultaneously minimizing the homeowner costs under the Time of Use rate structure. Based on these findings, we present several recommendations to improve the effectiveness of future load-shaping pilots and programs.

Introduction

Until recently, most US buildings have had a passive role within the electrical grid, consuming around 75% of electricity and driving daily peak power demand (DOE 2020). However, many electrical loads in buildings can be flexible and, if controlled in a coordinated way to draw different levels of electricity at specific times, could potentially support the electric grid, becoming interactive elements of the grid (DOE 2020).

Currently, building grid services (e.g., peak demand reduction) are still limited as they are usually called infrequently and over short periods of time (typically during extreme weather), and using a relatively small fraction of customers providing grid services. Expanding the use and flexibility of these services in the near future could help improve grid stability and reduce costs due to factors including the growth of electric demand and intermittent renewable electricity generation and more frequent extreme weather (Roth 2019).

Several theoretical studies and field pilots researched the potential for continuous building grid services. For example, the SunDial project designed, developed, and implemented an integrated supervisory control system. Over the course of an 18-month pilot comprising three commercial and industrial buildings (~3.8MW, peak), a 1.5 MW solar PV farm, and a 0.5MW/1.0 MWh battery energy storage system (BESS), the supervisory control system optimally dispatched facility loads and the BESS to mitigate potential problems from PV intermittency and power backfeed, as well as large ramps in grid power demand (Kromer, Roth, and Yip 2020). NREL's Foresee project considered a community of residential homes (Jin et al. 2017). That study included several model "smart homes" equipped with sensors that submetered power consumption of selected appliances (e.g., HVAC and electric domestic hot water (DHW) and performed model-predictive control (MPC) using simulations.

The "Packetized Energy" load management approach uses energy packetization and randomization to shape the aggregated electric demand. That is, the supervisory controller either accepts or denies requests to consume a "packet" of energy coming from flexible loads that range from EV charging stations (EVCS) and BESS to thermostatically controlled devices (i.e., AC and DHW) (Almassalkhi, Frolik and Hines 2022). Whereas this approach demonstrates scalability (St. John 2024), to provide *continuous* grid services it still needs to overcome several challenges, such as the physical constraint on how frequently controlled appliances can switch on and off (which can degrade HVAC equipment and dehumidification performance) and the need to accurately forecast HVAC loads and their flexibility beyond a basic (e.g., precool, ride-through, and recovery) demand response event.

While many homes could be a significant grid resource, there are several barriers to converting them into grid-interactive buildings and perform optimal load scheduling. Electricity consumption data are required to model the appliance power consumption and its flexibility. Unlike in NREL's "smart homes," acquiring data for individual loads (i.e., appliances) in existing homes typically implies installation of sensors to submeter these loads, which is inconvenient, costly, and sometimes disruptive to the homeowners. Another option, installing new connected devices is costly unless appliances are already at end-of-life.

For HVAC loads in MPC applications, the underlying building thermal model (e.g., first order greybox model as in Jin et al. 2017) and its calibration methods are usually too coarse to adequately represent thermal behavior of homes under extreme weather conditions, even as those are the conditions most relevant to load shedding or shifting grid services. Lin, Middelkoop and Barooah (2012) report a breakdown of a greybox model calibrated on August data and applied to October data for an office building. The discrepancy between the predicted and actual indoor temperatures (T_{in}) was up to 13°F. Using more accurate models such as EnergyPlus (~200 parameters) or higher-order greybox models coupled with semi-automatic calibration tools (Sturzenegger et al. 2012) is not possible scalable in many practical applications either because no detailed home information is available for automated model calibration. Moreover, nonlinear effects (such as moisture transport and AC performance curves) are usually neglected in MPC modeling as well as air leakage (Li et al. 2021).

Lastly, conventional MPC approaches are not directly applicable to a community of homes as computationally they typically have a polynomial complexity with the number of homes/devices to control, rendering the control scheme computationally expensive. Moreover, such grid services as load shaping may require a computational time step of one minute or less, which is much smaller than those usually implemented (e.g., 30-minute time step and 8-hour MPC horizon in Jin et al, 2017). Finally, the underlying control problem is not necessarily convex thus precluding the use of conventional gradient-based optimization schemes.

We have recently completed a DOE-supported project (Zeifman et al. 2025) that attempts to overcome these barriers. We are using homes equipped with the Sense Home Energy Monitor (HEM) installed on the home's main breaker that nonintrusively disaggregates the whole-home power signal into the power time series of individual major appliances. Thanks to Sense's agreement with a CT vendor, Sense also obtains CT data in some customers' homes, thus providing two independent data streams (power draw of major appliances and thermostat setpoints, room temperature and relative humidity [RH], and HVAC state and runtime). Using these two data streams, concurrent weather data, and basic home metadata (floorspace and ZIP code), we developed and validated a third-order greybox thermal model for homes with a central forced-air HVAC system with a gas furnace and vapor-compression cooling cycle (typical for a significant portion of U.S. homes), as well as a method for its automatic calibration. The model provides superior predictive accuracy and over longer time horizons than the state-of-the-art models (Zeifman and Roth 2026). In addition, our partners at Boston University (BU) developed optimization algorithms offering almost linear computational complexity with regard to the number of participating households and controllable devices (Ozcan 2025).

This paper presents a field pilot and virtual pilot conducted as part of the DOE-supported project. We developed and implemented a program to test the developed models and optimization algorithms in the field, i.e., on occupied homes. Section 1 presents the pilot design, including the recruitment process and technical means for conducting the pilot. In Section 2, we present and explain the pilot results. Given the relatively small size of the recruited communities, we enhanced the pilot's scope by performing Monte Carlo (MC) simulations with semi-synthetic data for a larger community size. In Section 3, we present the design and results of the virtual pilot and conclude the paper with a summary.

Pilot Design

Given the relatively low spatial density of Sense customers and low expected recruitment rates, we decided to create two communities, in the Northeast and Pacific Northwest, each composed of relatively co-located homes. To better represent a community with modern (or future) grid-interactive homes, we complemented these "core" homes with data from additional homes with EV(s) and/or PV and use only the underlying EV data from such homes to artificially boost the EV penetration rates.[1]

Recruitment

We imposed the following eligibility criteria for candidate homes:

[1] This assumes that the presence of EVs and solar PV is largely independent of building thermal response and HVAC power draw.

- Two recruitment regions: Pacific Northwest (pool of ~5,400 Sense customers circa 2023) and Northeast (~7,460 Sense customers circa 2023).
- The initial requirements for this pilot are that "core" homes in each area must be:
    - within the same climate zone,
    - within the same time zone,
    - with an active Sense monitor and at least one year of data, and
    - within a 200-mile radius.
- Additional "core" home characteristics specific to our modeling approach:
    - Must be a single-family home,
    - Must use a central HVAC unit, with a furnace (gas- or oil- fueled) for space heating, and central AC for cooling,
    - Must have no more than two HVAC zones (i.e., no more than two CTs).

Sense sent personalized emails to potential pilot candidates in two recruitment waves, detailing what the pilot is about and how it can help the customer to support the grid and reduce GHG emissions. To increase the pool of potential participants, for the second wave we dropped the furnace requirement and recruited homes that had a central AC for cooling, regardless to their heating system (because of the lack of heating data, it was not possible to calibrate our thermal model using the previously developed approach (Zeifman and Roth 2026).

Data handling tools for the pilot

We used a combination of tools and data providers to execute this pilot:

- *Retool:* It fulfills the function of frontend and backend (including the database). The code is written mostly in JavaScript and the database is SQL. Retool also provides email notification support.
- *Webhook:* Powered by Retool, this listener triggers when a device's status changes.
- *Enode:* Provider of the vehicle and HVAC data. Enode has public APIs that we used to retrieve the device's information. Also provides a UI for the device's registration by generating a custom link for each user.
- *Typeform:* Platform that hosts our survey. When a user completed the survey, they were redirected to the Retool application.
- *Sense API*: The Sense API obtained data from each home's HEM (Status and kW).

These tools can obtain data from and issue dispatch command to CTs and EV chargers from multiple vendors, but we found that the number of homes with electric compared to gas water heaters was small, and transforming those water heaters into controllable ones would require purchasing and installing extra hardware that would not have been possible with the personnel resources and timeline for this pilot. Therefore, we had to exclude DHW as a flexible load from the pilot.

Finalizing the Pilot Communities

We verified that the connected devices in candidate homes behaved as expected, i.e., that during a heat wave CT(s) reported "cooling" status, and reasonable indoor temperature T_{in} and T_{set} values We also checked for consistency between the disaggregated AC and CT data.

Similarly, we inspected the most recent data on EV homes. We then added all homes with validated controllable loads to the pilot to construct two communities of candidate homes, one in the Northeast and another in the Northwest with the following distribution of characteristics:

- Northeast: 15[2] "core" homes (five with two CTs, six with PV, two with EVs) + data from eight additional homes with EVs
- Northwest: 15 "core" homes (zero with two CTs, five with PV, two with EV) + data from six additional homes with EVs

Pilot load shaping goal

In the pilot, we wanted to perform load shaping on days with significant AC usage, i.e., days with forecasted hot weather. In particular, we wanted to flatten the aggregated community load $Q(t)$ as much as possible during the event hours, typically 10 AM to 6 PM, to mitigate the possibility of negative power flow due to the ~30% PV penetration:

$$Q(t) = \sum_i^N W_{HVAC}^i(t) + W_{EV}^i(t) + W_{other}^i(t) - W_{PV}^i(t), \tag{1}$$

In Eq. (1), $W_{\mathrm{HVAC}}(t)$, $W_{\mathrm{EV}}(t)$, $W_{\mathrm{other}}(t)$, and $W_{\mathrm{PV}}(t)$ are HVAC power draw, EV charging, non-HVAC and non-EV (i.e., all other electric devices) power draw, and PV power generation, respectively, of the *i*-th home of *N* homes, and *t* is the time at 15-minute intervals.

Optimization and trajectory generation

Given the nonlinearity of our HVAC model, the distributed-optimization MPC method developed by our BU partners (Ozcan 2025) was modified to work with *trajectories*, i.e., potential power consumption profiles of individual homes. A branch-and-cut type of algorithm then finds an optimal combination of trajectories (in the sense of Q(t) flattening) out of many candidate trajectories and provides an optimally selected trajectory for each home. Since Retool only allows for manually changing temperature setpoint profiles, we decided to use a once-a-day optimization rather than continuous MPC.

However, the optimization tool developed by the BU team appeared to output trivial solutions (e.g., zero HVAC usage for a majority of homes) in initial testing. With very limited time for debugging, we implemented a naïve optimization routine for the Pilot:

- For each home, select a candidate HVAC trajectory at random
- Calculate the community Q-vector according to Eqn. (1) using the selected trajectories
- Repeat steps 1 and 2 many (~10,000) times
- Among the generated Q-vectors, select the one with the lowest mean value and flattest profile (determined by calculating standard deviation) during the event period and take the associated HVAC trajectories as optimal.

Generating HVAC power trajectories

We designed several types of trajectories: "offset"; "simple moving window"; "simple breakpoint"; "symmetric"; and "pre-cooling". With the exception of the "offset" type, these

[2] A CT in one home stopped working in mid-June, so we had to exclude that home from the pilot

trajectories utilize the base setpoint profile $T_0(t)$ that we derived from each home's historical data.

- **Offset**: Rather than prescribing a specific T_{set}, T_{set} changes by a specified offset relative to the T_{set} already scheduled by the household.
- **Simple moving window**: T_{set} equals T_0 at the beginning of the event, increased/decreased by 2-3 °F for 2-3 hours, then returned to $T_0(t)$ for the remainder of the event.
- **Simple breakpoint**: T_{set} equals T_0 at the beginning of the event, then increased/decreased by 2-3 °F after 2-3 hours until the end of the event.
- **Symmetric**: The $T_{set}(t)$ for the duration of the event is designed to be symmetric about $T_0(t)$. T_{set} is initially set to T_0 at the beginning of the event, then increased/decreased by 2-3 °F for 2-3 hours. T_{set} is reset to $T_0(t)$ for a period, then increased/decreased (opposite direction as before) for 2-3 hours, then reset to $T_0(t)$ until the end of the event.
- **Precooling:** At the beginning of the event, T_{set} decreased 2-3 °F from T_0 for 3-4 hours (precooling), then increased 2-3 °F above T_0 until the end of the event ("ride-through").

We generated 10-20 HVAC trajectories per "core" home; EV trajectories can be generated, e.g., by varying the charging start time and duration of charging subject to the initial and final charge limitation.

Predictive models

To perform load shaping for a community of homes, we need predictive models for:

- **AC loads**. We used our 3rd-order greybox model with 14 parameters to predict T_{in} and AC load over a time horizon (up to 72 hours), given the current T_{in} and the following vector inputs: weather forecast (T_{out}, wind speed, solar global horizontal irradiance [GHI] and elevation angle) and T_{set}. Whereas we developed an automated estimation technique for building thermal model calibration using heating season data, with AC performance curves estimated from cooling season data, we had to assume a constant performance curve for the recruited "core" homes since no indoor RH data were provided. Nonetheless, we had reasonable T_{in} predictions, with average root-mean-squared error (RMSE) over a 36-hour horizon of 0.55 °C and 0.79 °C for the Northeast and Northwest homes, respectively, for the limited historical data available (May-July 2024).
- **PV generation**. We are using a machine-learning method (Support Vector Machine Regression) with predictive variables GHI, solar azimuth and elevation angles, outside temperature and wind speed. The model was trained using ~one year of historical power generation data that were available for the Sense customers. The model accuracy was 16% and 17.7% in terms of RMSE over a 24-hour prediction horizon for the Northeast (6 PV homes) and Northwest (5 PV homes) communities respectively.
- **EV charging**. Our analysis of empirical EV data suggests that the EV charging profile is usually a square wave with a near-constant charging rate (3-11 kW depending on the EV/charging station combination). In addition to the kW and % charging rates (straightforward estimation from historical data), the EV's state of charge (SOC) at the start and end of charging are needed to define the charging process (practically all EVs in our sample stopped charging at 80%). We obtained those data from the EVCSs.

- **Non-HVAC loads**. Given potential correlations between these loads and weather conditions, we used statistical resampling to model these loads, analyzing historical weather and selecting the day within the past 60 days that most closely matches the forecasted weather in terms of cooling degree days and average GHI. For that day, we calculate the home's non-HVAC power consumption profile and use it as a predictor.

Pilot Events and Results

Table 1 summarizes the 11 events we ran in August and September 2024. For each event, we ran the models and optimization algorithm to solve for the dispatch schedule for all controlled loads on a day-ahead basis. Sense specified this approach so that customers could be notified of event times well ahead of time; however, this prevented us from adjusting dispatch schedules as the actual daily load shape developed and, in particular, T_{out} and GHI weather forecasts were refined.

Table 1. Summary of pilot events

Event	Date	Region	Devices	Time	# Homes	Avg. T_{out} over HVAC Event	PV	Opt-outs
1	24 Jul	NE	EV	9am 2pm	10		–	2
2	25 Jul	NW	EV	9am – 2pm	8		–	3
3	08 Aug	NW	EV	9am – 2pm	8		–	1
4	08 Aug	NW	HVAC	9am – 6pm	11	27.1 ºC	4	3
5	14 Aug	NE	EV	8am – 2pm	10		–	4
6	14 Aug	NE	HVAC	9am – 5pm	12	25.5	5	2
7	28 Aug	NE	HVAC	9am – 5pm	11	28.9	4	4
8	30 Aug	NW	HVAC	12pm – 6pm	13	29.8	4	2
9	05 Sep	NW	HVAC	9am – 6pm	10	30.1	4	0
10	06 Sep	NW	HVAC	9am – 6pm	10	30.9	4	2
11	13 Sep	NE	HVAC	9am – 6pm	10	25.4	4	1

Our first events were simple EV events. As have others, we encountered several limitations with managing EV loads. First, for any command to shift EV charging to have an impact, an EV must be connected to the charger. Second, as soon as an EV is connected to a charger, it typically starts charging instantly. Consequently, if we want an EV to charge during a specific time window, we often need to issue several "Stop charging" commands prior to the specified charging time. We tried using a "Cascading Events" scheme (regular issuing of "Stop charging" command), but the implementation was unsuccessful, crashing Retool. Finally, through direct engagement with users, the Sense team identified connectivity issues, particularly for EVs parked in garages with poor internet connectivity. This hindered effective execution of events, as instructions could not reliably reach the EVs in a timely manner.

After reconfiguring the Retool application, we attempted two additional EV events. Although they did not cause Retool to crash, the EVCSs still did not follow the specified charging commands. First, we modified the notification text and asked EV owners to connect their EV either the evening before (9 pm) or the morning of the event (9 am), and designed the cascading commands accordingly. This prevented error messages and, indeed, most users did connect their EVs as requested, but we found another challenge: most connected EVs had about

70-80% SOC at the time of connection, leaving very little opportunity for further charging. Finally, we observed that many users disconnected their EVs and apparently left in the late morning hours.

Overall, only one or two EVs completed the desired event charging profile (out of 8-10 homes participating). These three challenges, i.e., few EVs plugged into the EVCS, high SOC of plugged in EVs; and EVs disconnecting during events, make EV charging unreliable for the Pilot, constraining their load shifting potential. With more advanced EVCS communication available, we would have addressed the first two challenges by issuing Stop Charging command as soon as the user connects their EV to the charger, as opposed to the after 9 pm + 1-hour repeats in the cascading design. However, we could not control when people used (i.e., disconnected) their EVs, which significantly limits the value of home EVCS control during much of the day. Thus, we only controlled HVAC during the remaining events.

Predictive model performance

Given the limitations of the Enode platform (e.g., did not report HVAC runtime, integer only values of temperature, no RH) and inevitable errors in the disaggregated AC power time series, we assessed the results of the thermal model performance using the temperature prediction accuracy rather than HVAC runtime.

For each event, we estimated the average goodness-of-fit (GOF) for each home's thermal model, calculated as the RMSE between forecasted and actual T_{in} over the event. Table 2 compiles GOF values for different events (dashes indicate a home did not participate in an event). For each community, GOF values greater than one standard deviation (σ) above the mean (μ) are highlighted (Northeast homes have an average GOF of 0.7 °C, with a standard deviation of 0.4 °C; Northwest homes have an average of 0.5 °C, with a standard deviation of 0.2 °C). In general, the GOF values fall within the range of the calibration GOF values. We found negligible correlation between GOF and outdoor temperature (Tout).

Table 2. Individual home thermal model goodness of fit; values in °C. A dash ("–") indicates a home that didn't participate in that event. Highlighted cells are greater than $\mu + 1\sigma$.

Home	NE	NW	14-Aug	28-Aug	13-Sep	08-Aug	30-Aug	05-Sep	06-Sep	Average	Calibration
h222165	X		0.5	0.6	0.8					0.6	0.4
h1759	X		0.3	0.5	0.7					0.5	0.3
h29351	X		0.5	1.8	0.6					1.0	n/a
h282167	X		–	0.6	0.5					0.6	0.9
h209478	X		0.4	0.6	0.3					0.4	1.1
h249271	X		1.7	1.0	0.6					1.1	0.6
h211568	X		0.4	0.7	1.2					0.8	0.5
h231859	X		0.5	0.4	–					0.5	0.3
h272658	X		1.3	0.8	0.4					0.8	0.7
h286768	X		0.8	0.6	1.4					0.9	0.4
h199310	X		0.6	0.4	0.3					0.4	0.5
h8134		X				0.4	0.6	0.5	0.8	0.6	1.0
h220799		X				–	0.3	–	–	0.3	1.5
h292306		X				0.4	0.4	0.3	0.3	0.4	0.5
h196828		X				0.5	0.5	1.0	1.2	0.8	1.0
h206113		X				–	0.3	0.4	0.6	0.4	0.9
h281191		X				0.7	0.5	0.5	0.7	0.6	0.5

h265162		X				–	0.4	–	–	0.4	n/a
h272697		X				–	0.6	0.4	0.5	0.5	0.5
h285249		X				0.3	0.5	0.6	0.6	0.5	0.6
h38559		X				0.6	0.5	0.8	0.8	0.7	0.8
h33466		X				–	0.5	–	–	0.5	n/a
h276078		X				0.3	0.4	0.3	0.3	0.3	0.8
h183492		X				0.4	0.4	0.5	0.7	0.5	0.6
T_{out} (ºC, mean)			25.5	28.9	25.4	27.1	29.8	30.1	30.9		
T_{out} (ºC, max)			27.1	30.2	27.5	31.5	31.8	35.2	34.4		

We also evaluated the performance of the hourly PV prediction models used in planning events. Table 3 displays RMSE of aggregate PV generation predictions for each event, both in kilowatts and as a percentage of maximum generation. These predictions were made based on weather forecasts obtained from our vendor Solcast.

August 28 had significant morning cloud cover, contributing to a higher PV prediction error. On August 14, one of the three homes with PV did not produce power until the afternoon, leading to a higher error. With the exception of these two outliers, the aggregate PV prediction error was within 15% of the maximum daily generation, which is similar to the model training accuracy. For comparison, we estimated that the differences between day-ahead Tout and GHI predictions during the events an event were around 1.5 ºC and 7%, respectively.

Table 3 PV aggregate prediction performance.

Event	Region	RMSE (kW)	Maximum (kW)	Percent Error (%)
08-Aug	NW	1.1	13.4	10.8
14-Aug	NE	5.8	20.1	28.8
28-Aug	NE	6.3	22.8	27.4
30-Aug	NW	1.8	16.4	10.8
05-Sep	NW	1.1	15.2	7.1
06-Sep	NW	1.9	14.5	13.2
13-Sep	NE	1.7	23.9	7.0

Load shaping results

Figures 1 and 2 show the load shaping results for four events, two each with the Northeast and Northwest communities. Each plot shows three net power lines as follows:

- Baseline = HVAC (regression-based prediction) + NonHVAC (“everything else”, sampled from the historical data) – PV (forecast)
- Prediction = HVAC (optimized) + NonHVAC (historical) – PV (forecast)
- Recorded = Mains (measured, actual) – PV (measured, actual)

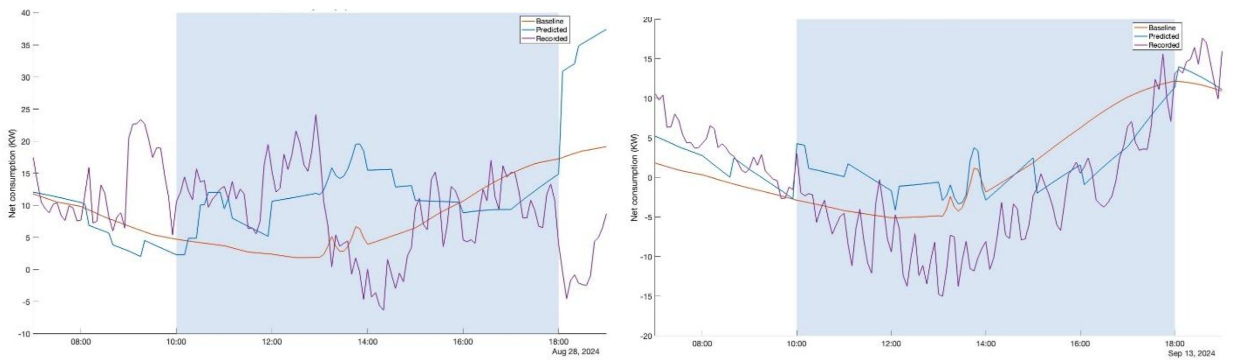

Figure 1. Results for two Northeast Events (28 August and 13 September)

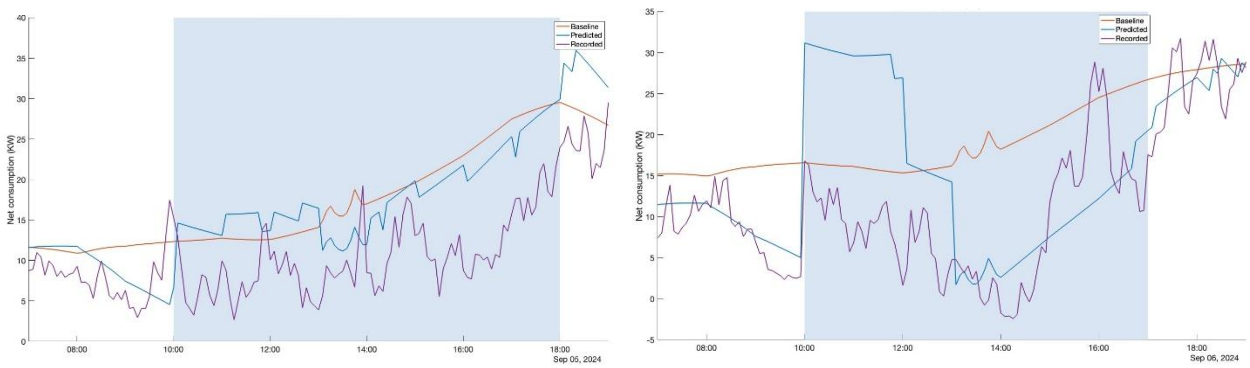

Figure 2. Results for two Norwest events (5 and 6 September)

Overall, the plots indicate that load shifting did occur during the event window (10 am – 6 pm), although the actual load did not appear to be significantly flatter than the baseline. Since Tables 2 and 3 indicate the reasonable accuracy of the predictive models in this field testing, we investigated potential reasons for this inconclusive load shaping result. Specifically, we think significant improvements in load shaping can be achieved by:

- Increasing the number of participating homes;
- Applying an MPC scheme that continuously updates modeling and control actions rather than the single ~day-ahead prediction we had to use in the pilot implementation;
- Implementing the corrected BU optimization approach rather than the inefficient optimization scheme;
- More accurate AC power disaggregation across homes for model calibration;
- More robust forecasting techniques for non-HVAC loads; and
- Using a data integration that preserves more information, notably HVAC runtime and RH and does not round the T_{in} reading (as Enode did).

Virtual Pilot with Semi-synthetic Data

Due to the significant pilot limitations noted above, we enhanced the pilot scope by performing Monte Carlo (MC) simulations with semi-synthetic data for a larger community size

(100 to 1,000 homes) to test the advantages of the models and optimization algorithms. This enabled us to model how realistic communities of homes with varying PV and EV penetrations would perform in typical DR programs with and without optimization in terms of community-level load shaping and individual home savings. Moreover, we could also study the effect of advanced HVAC and PV models, i.e., our modeling vs apparently coarser models (e.g., linear models for HVAC, regression models for PV).

We developed a MC procedure to “create” a virtual but realistic community of homes, based on the data collected over the field Pilot period and applying our models and optimization algorithms. The procedure builds upon the “core” homes, i.e., the homes we used for AC modeling and control, and adds to these homes PV systems and/or EVs at random, according to pre-defined rates of PV- and EV-penetrations.

While preparing for this virtual pilot, we realized there might be correlations between HVAC loads, everything-else loads, T_{set}, and temperature prediction errors. Accordingly, initially we just resampled the core homes, adding PV and/or EVs to them based on PV system characteristics and charging patterns sampled from other field pilot homes. In the future, we could model more advanced scenarios, e.g., using an empirical distribution of home floorspace for an actual community (e.g., for a given ZIP code), or a joint distribution of floorspace, PV size, and EV charging rate.

The basic sampling procedure for defining the characteristics and load shapes for each home in the virtual pilot was to:

1. Select either a community in the Northeast (NE) or Northwest (NW) to simulate.
2. Select the most frequent ZIP code for a selected community and obtain weather data.
3. Obtain 24-h T_{set} profiles for each “core” home and for each day during the summer of 2024. For each obtained T_{set} profile, record the date and indicate weekday or weekend.
4. Obtain thermal model parameters (taken from the field pilot), calculate T_{in} predictions over the 24-h period and calculate the corresponding indoor T_{in} prediction error profile, for each day during the summer of 2024. For each obtained T_{set} profile also record the date.
5. Obtain PV prediction parameter sets (one set per PV home, taken from the field pilot) for each community. Record square footage for each set. Calculate 24-h PV prediction error profiles for each day during the summer of 2024, also indicating dates (similar to #4).
6. Calculate EV characteristic set (charging rate in %/h and kW, initial and final SOC) for each community (one set per EV, taken from the field pilot).
7. Obtain 24-h power profiles for “everything else” for each core home and for each day during the summer of 2024. For each error profile also record date.

We then applied a somewhat similar process to generate a community of homes for each MC virtual pilot (see Zeifman et al. 2025 for the specific steps of that process).

Figure 3 shows the results of the MC simulation of a community of homes based on the Northeast and Northwest original communities. The “duck-curve” shape is clearly seen in both plots (the original community PV penetration rates are preserved in MC simulations).

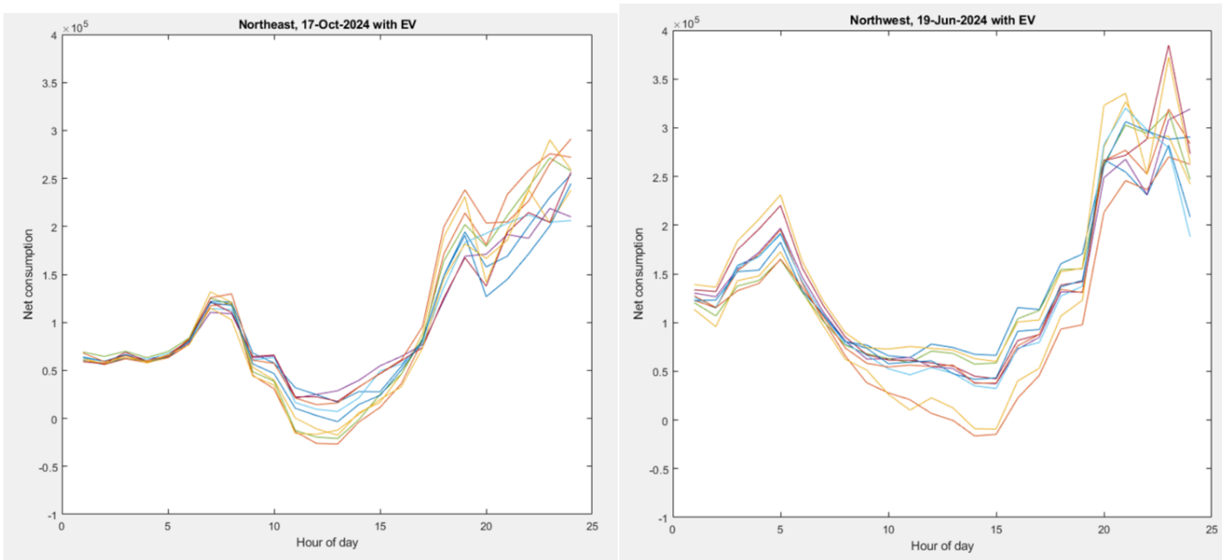

Figure 3. Net power profiles of a MC community of 100 homes in NE with 10 repetitions; each color trace represents the results of a repetition. The Y axis units are Watts times 105. Left: Northeast, 17 October 2024. Right: Northwest, 19 June 2024.

Virtual Pilot Results

Figure 4 shows optimization results (flattening of the net power) for various sizes of the simulated community. Only AC usage has been optimized, and the overall net power is being optimized for the time from 8 am to 8 pm (note that this time window exceeds a typical DR time window of three or four hours by a factor of three or four). It can be seen that the variability along the targeted flat line decreases with community size, and that the optimal curve achieved significantly differs from the baseline.

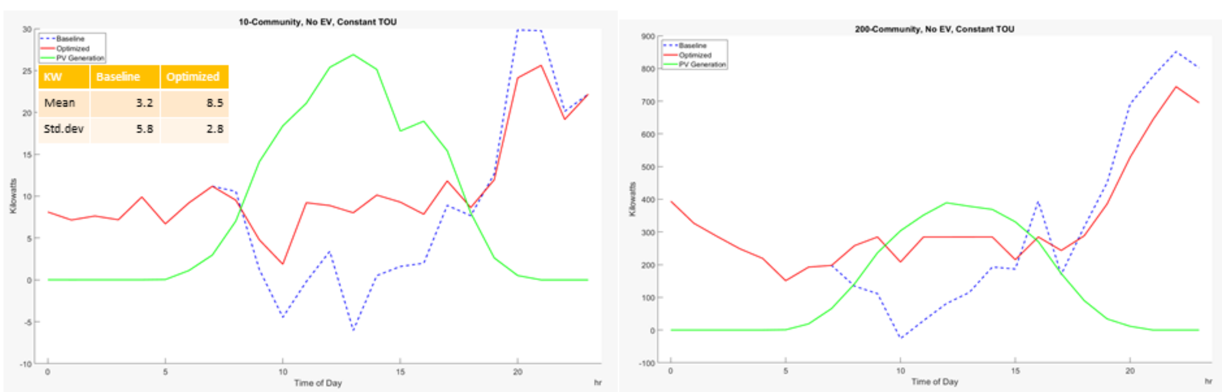

Figure 4. Flattening the net power consumption for a MC community of homes from 8 am to 8 pm by optimal selection of AC Tset trajectories only (no EV control) for each home, BU optimization algorithm is used. Community size ranges from 10 (left) to 200 (right) homes.

If we can control both AC and a sizeable population of EVs connected during events that can accept additional charging, the resultant optimized curve becomes essentially flat (see Figure 5). Note that for this simulation, we assumed a uniform availability of EVs to charge in time.

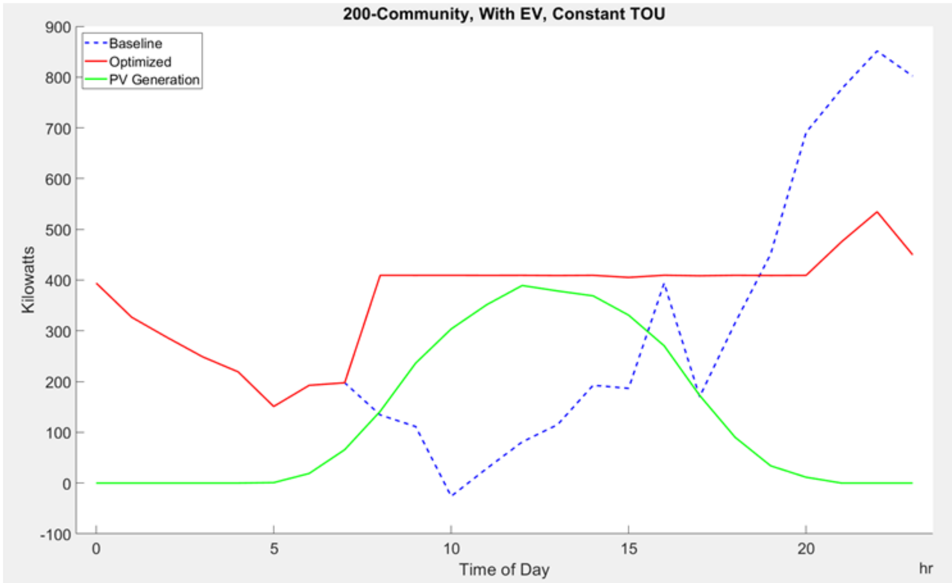

Figure 5. Flattening the net power profile of a community of homes from 8 am to 8 pm by optimal selection of AC Tset trajectories and EV charging for each home. Community size is 200 homes.

Next, we evaluated how difficult it is to select the optimal trajectories using the BU-developed algorithms as compared to the naïve optimization used for the field pilot (see Section 2). Figure 6 shows the community net load profiles calculated for a random sample of generated AC trajectories vs the optimal solution. Given that we generate about 100 trajectories for each of the 200 homes in the community, the total number of potential trajectories (100^{200}) is too large to show all the trajectories. Since the shown sample is representative, the vast majority of the trajectories would be within the grey-colored domain in the Figure, whereas the optimal solution is located far away from them. This implies that typical optimization schemes used in MPC applications for energy (not only our naïve optimization approach, but also the approach being used in Foresee (Jin et al. 2017)) would struggle to deliver the optimal solution.

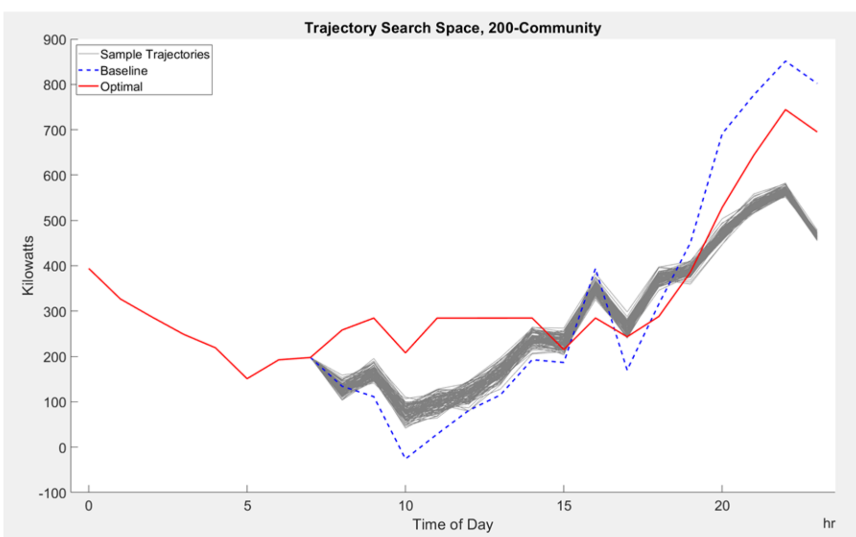

Figure 6. Net community loads calculated for a random sample of generated AC trajectories vs the optimal solution. Community size is 200 homes.

The BU algorithm can also concurrently optimize for two objectives: aggregate load shaping and individual home cost minimization assuming time-of-use (TOU) pricing. Figures 7 and 8 show the results for the case with TOU doubled from 4-10 pm, using only AC flexibility. The flatted curve indicates the optimization algorithms selected AC setpoint trajectories yielding near-optimal performance.

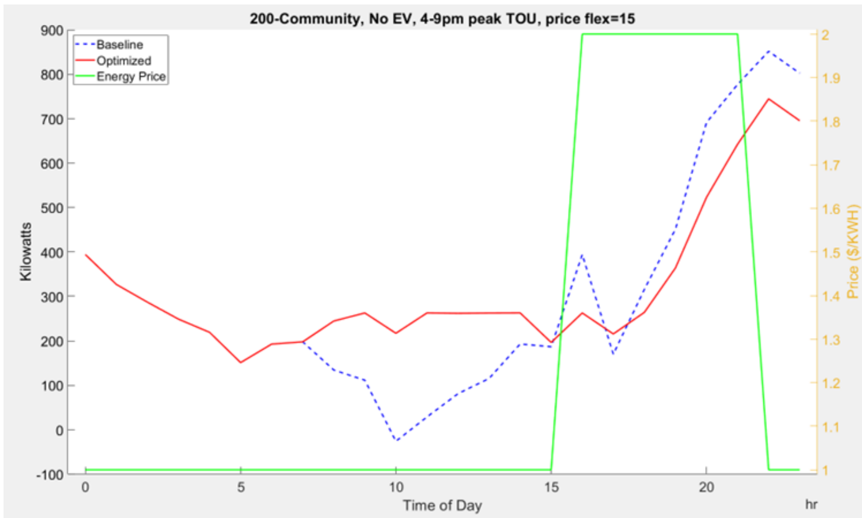

Figure 7. Overall community power profile under simultaneous TOU and community load flattening optimization. Community size is 200 homes.

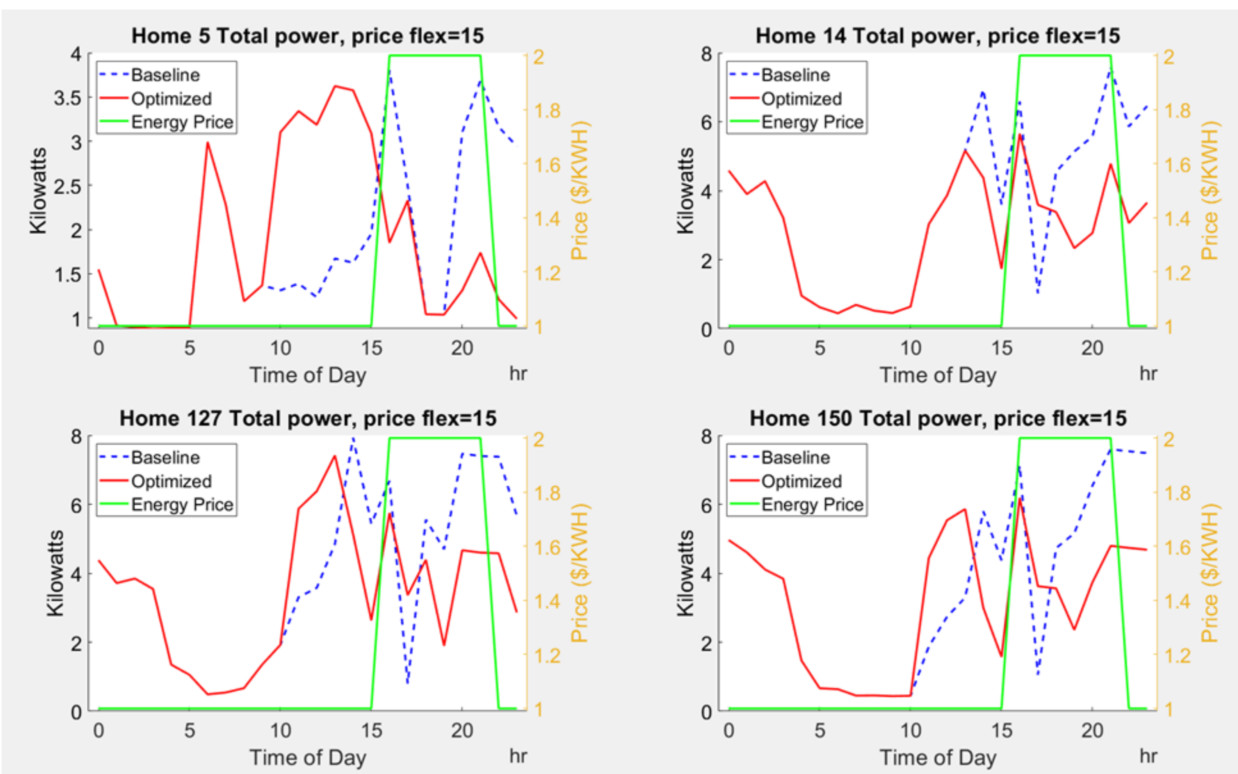

Figure 8. Optimization of individual home power consumption by shifting AC setpoints under TOU pricing AND flattening the community net load for homes shown in Figure 8.

Conclusions and Recommendations

The main goals of the pilot were to test the performance of the predictive models and optimization algorithms, and to gain experience with real-world implementation. We learned several things from the field pilot despite its small sample size:

- The predictive models for AC loads performed well, predicting T_{in} on days with HVAC load-shifting events with an average RMSE of 0.5 to 0.7 °C (predictions made at start of the event day, i.e., at midnight).
- The aggregate PV prediction error was within 15% of the maximum daily generation on event days, except one event that had appreciable cloud cover (challenging forecasting conditions) and another when a home did not generate PV for part of the day.
- The override rate was low for CTs but a large portion of EVs disconnected during events (presumably to drive the EV).
- EV charging has a much lower realizable demand management potential than typical EV charging levels because a large number of EVs are not connected during events, have a high SOC when connected, and a sizeable fraction disconnect during events.
- A variety of factors can compromise the ability to effectively implement load shaping functionality, including (but certainly not limited to):
    - Using an optimization scheme that cannot solve for the near-optimal dispatch schedule;
    - Inability to regularly update control dispatch signals throughout the day based upon actual power generation and consumption profiles;
    - Errors in AC power disaggregation;
    - Less robust forecasting techniques for non-HVAC loads; and
    - Data platforms that provide more limited inputs for model calibration and, as a result, degrade model fidelity.

Our virtual pilot using semi-synthetic data sampled from the pilot homes clearly demonstrated the value of a robust, computationally-efficient optimization algorithm for identifying optimal/near-optimal dispatch schedules for a community of homes. That is, the typical optimization algorithms underlying MPC methods perform significantly worse than approach developed by our team, as their much lower computational efficiency effectively precludes finding a near-optimal solution for a sizeable number of homes/devices. The optimization approach we applied can effectively flatten the aggregate load profile over a 12-hour time frame, which is appreciably longer than typical DR events. In addition – and not surprisingly – we showed that managing EV charging can dramatically enhance load shaping potential because they are a very flexible and predictable load under a wide range of conditions.

Our recommendations for future load shaping pilots follow directly from the discussions above, i.e., to: 1) implement more advanced data-driven models to improve prediction accuracy for key flexible loads; 2) use efficient optimization algorithms to enable selection of a set of near-optimal dispatch schedules for controllable loads; and 3) develop program designs for local load shaping applications that resolve the tension between customers who prefer day-ahead notification of events and the ability to refine dispatch schedules as more information becomes available.

Acknowledgements

This work was supported in part by the U.S. Department of Energy Building Technologies Office (BTO) under Contract #DE-EE 0009696. We would like to extend our gratitude to Wyatt Merrill, DOE-BTO Technology Manager and Coriana Fitz, DOE Technical Project Officer for their support and fruitful discussions. We would like to thank Jeremiah Miller and Marc El Hachem for their work and meaningful discussions during the early stages of Pilot preparation, and Matthew Kromer for his contributions.

References

Almassalkhi, M., Frolik, J., and Hines, P. 2022 "Packetizing the Power Grid: The rules of the Internet can also Balance Electricity Supply and Demand," IEEE Spectrum 59, pp. 42-47.

DOE (US Department of Energy). 2020."Grid-Integrative Efficient Buildings." https://www.energy.gov/sites/default/files/2020/09/f79/bto-geb-project-summary-093020.pdf

Jin X. et al. 2017. "Foresee: A user-centric home energy management system for energy efficiency and demand response," Applied Energy 205, pp. 1583-1595.

Kromer, M, Roth, K, Yip, T. 2020. "Optimizing DER Dispatch in a Renewables Dominant Distribution Network Using a Virtual Power Plant." 2020 IEEE (PESGM).

Li, Y. et al. 2021, "Grey-box modeling and application for building energy simulations – A critical review," Renewable and Sustainable Energy Reviews 146, 111174.

Lin, Y., T. Middelkoop, P. Barooah. 2012. "Issues in identification of control-oriented thermal models of zones in multi-zone buildings," IEEE Conference on Decision and Control.

Ozcan, E. C. 2025. Distributed Solution Techniques to Regulate the Load Consumption of a Residential Neighborhood. PhD thesis, Boston University.

Roth, A. 2019. "Grid-interactive Efficient Buildings Technical Report Series," DOE, https://www1.eere.energy.gov/buildings/pdfs/75478.pdf

St. John, J. 2024 "Fine-tuning how homes can help the grid as 'virtual power plants'" Canary Media. https://www.canarymedia.com/articles/virtual-power-plants/fine-tuning-how-homes-can-help-the-grid-as-virtual-power-plants

Sturzenegger D. et al. 2012. "Semi-Automated Modular Modeling of Buildings for Model Predictive Control," BuildSys Conference Proceedings, pp. 99-106.

Zeifman, M. and Roth, K. 2026 "A practical thermal model of residential buildings to enable grid interactivity." Submitted.

Zeifman, M. et al. 2025 "High-fidelity self-learning tool for residential load and load flexibility forecasting," Final Technical Report, DOE, available at https://www.osti.gov/biblio/3001405