

Almost DIY: A nonintrusive way to integrate homes with the electric grid

Michael Zeifman, Fraunhofer USA Center for Manufacturing Innovation

Hannah Fontenot, Fraunhofer USA Center for Manufacturing Innovation

Kurt Roth, Fraunhofer USA Center for Manufacturing Innovation

Donna Leclerc, Sense Corporation

Scott Taylor, Sense Corporation

George Zavaliagkos, Sense Corporation

Erhan Can Ozcan, Boston University

Ioannis Paschalidis, Boston University

ABSTRACT

The concept of grid-integrated residential buildings usually implies new homes equipped with sensory and actuation technologies along with energy-efficient appliances and electricity storage that enable optimal scheduling of flexible electric loads to provide grid services. Currently, such technologically advanced homes represent a tiny fraction of US housing stock, and electricity storage is prohibitively expensive in most jurisdictions. To address this discrepancy, we consider a possibility of converting a portfolio of existing homes into a community of grid-interactive homes.

The proposed modeling concept utilizes two ubiquitous sensors, a nonintrusive energy monitor and a communicating thermostat to obtain the disaggregated appliance power and communicating thermostat data streams along with weather-station data to automatically train a hybrid grey-box/machine-learning model to 1) accurately monitor and forecast device loads and 2) perform model-predictive control (MPC)-based load management. Throughout each day, the model will monitor and forecast major loads (HVAC, water heating, user-controlled appliances) and their flexibility, as well as on-site PV power generation, to determine the optimal dispatch schedule of flexible loads that minimizes home energy costs while providing grid services under different use cases and tariff structures and maintaining occupant comfort.

In this paper, we present results of a first phase of the DOE-supported project. Specifically, we use heating-season data from twenty recruited homes to automatically calibrate a grey-box model for thermal load prediction, and we use a hybrid physics-based/ machine-learning approach to predict home PV generation. A novel inverse optimization method is proposed to optimally schedule home appliances to perform grid services.

Introduction

Until recently, buildings had a passive role within the electrical grid in the US, consuming a lion share of the electricity and also driving daily peak power demand [DOE, 2020]. However, many of electrical loads in buildings can be flexible, and, if realized in concert to draw electricity at specific times and at different levels, could potentially be used to increase the flexibility of demand. In this scenario, the buildings become interactive elements of the grid [DOE, 2020].

Currently, building grid services (e.g., peak demand reduction) are still limited as they are invoked infrequently and over short periods of time (typically during extreme weather), and are performed by a relatively small fraction of grid customers. These services need to be enhanced in the near future to address such factors as the growth of 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, in the SunDial project, an integrated system was designed, developed and demonstrated. The system included three larger commercial and industrial buildings (~3.8MW, peak), a 1.5 MW PV farm, and a 0.5MW/1.0 MWh electric energy storage system (ESS), with a supervisory control system that optimally managed facility loads and ESS charging and discharging to mitigate potential problems from PV intermittency and backfeed, as well as large ramps in grid power demand [Kromer, Roth, Yip, 2020]. A community of residential homes was considered by Jin et al. in NREL's Foresee project [Jin et al, 2017]. This study included several NREL model "smart homes" equipped with sensors to submeter power consumption on selected appliances (e.g., HVAC and DHW) and performed model-predictive control (MPC) using simulations. Again, an ESS was a vital component of the overall concept.

The battery storage is still expensive and could be inconvenient to homeowners/occupants nowadays [Battery Storage, 2022]. Existing residential homes with no ESS could be a significant resource for the grid services, but currently there are several barriers to convert existing homes into grid-integrated buildings and perform MPC-type of optimal load scheduling.

First, unlike in NREL's "smart homes", getting data on individual loads (i.e., appliances) in existing homes typically implies installation of sensors to submeter these loads, which is inconvenient and at times disruptive to the homeowners. Electricity consumption data are required to model the appliance power consumption and its flexibility.

Second, for HVAC loads in MPC applications, the underlying building thermal model (first order grey-box 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, but those are the conditions most relevant to load shedding or shifting grid services. Lin et al [2012] report a breakdown of a greybox model calibrated on August data and applied to October data in an office building. The discrepancy between the predicted and actual indoor temperatures was up to 13°F. Using more accurate models such as EnergyPlus (~200 parameters) or higher-order grey-box models coupled with semi-automatic tools [Sturzenegger et al, 2012] is not possible in our setting either because no detailed home information is available. Moreover, the nonlinear effects (such as moisture transport and AC performance curve) are usually neglected in MPC modeling as well as the air leakage.

Lastly, the conventional MPC approaches are not directly applicable to a community of homes. Many researchers have acknowledged that changing the electricity price by using naïve strategies may not provide desired results if the participants in grid services use the same optimization technique for scheduling. For example, simultaneous scheduling of flexible appliances towards the time slots with lower electricity prices leads to the rebound peaks [Li et al, 2012]. On the other hand, using different pricing models for different participants to achieve the overall desired grid effect implies that the dimensionality of the price-setting problem grows with the number of participants, rendering the scheme computationally expensive. Moreover, such grid services as load shaping may require a computational time step of one minute or less,

which is orders of magnitude smaller than those currently implemented (e.g., 30 minutes time step and 8-hour MPC horizon in [Jin et al, 2017]).

This paper presents our attempt to overcome these barriers. We are using homes equipped with the Sense [Sense, 2022] Home Energy Monitor (HEM) that is installed on the home's main breaker and nonintrusively disaggregates the overall power signal into the power time series of individual major appliances. Thanks to Sense's agreement with a CT vendor, Sense obtains CT data in many of their customers' homes, thus providing two independent data streams (power draw of major appliances and thermostat setpoints, room temperature, HVAC state and runtime). In addition to HVAC and electric water heaters (EWH) that are frequently used in Demand Response (DR) applications, we also incorporate Electric Vehicle (EV) charging as a third flexible load (a user-controlled appliance). In Section 1, we provide information on the homes and data we are using in this work. In Section 2, we detail our second-order greybox thermal model and the way we intend to make it self-calibrating using the available historical data from the homes. In Section 3, we outline a hybrid physics-based/ machine-learning approach to predict home PV generation. In Section 4, we explain our approach to optimally schedule the flexible loads in a community of homes. A computationally efficient Coordination Agent concurrently optimizes the performance of a portfolio of homes and individual homes in line with grid services required, subject to household preferences and grid constraints. Using a novel inverse optimization approach, the agent simultaneously and automatically converts several services (energy efficiency, load shifting, load shading and load shaping) into corresponding objective functions and has linear complexity over the number of homes and over the number of time steps (thus, allowing arbitrary time resolution).

1. Homes and Data

Out of Sense's extensive database of existing US customers, we selected twenty homes that have:

- A. central air conditioning (AC)
- B. Gas-fired furnace as the only major space heating device
- C. CTs
- D. EV charger
- E. Electric hot water heater (DHW)
- F. Rooftop solar (PV)
- G. At least a year of available data.

The requirements A, D, and E correspond to the presence of potentially flexible loads that we discussed in Introduction. Requirement B is specific to our thermal model (see next Section) and would be bypassed in future work. Requirements C and G correspond to data availability. Requirement F is natural for dealing with the scenario of high PV penetration. The Sense HEM can submeter PV generation; moreover, the PV generation data can inform the thermal model (see next Section).

The electric data for each house comprise 1-minute-resolution power consumption data on:

- Mains (overall power)
- PV generation
- Disaggregated data on:
 - AC

- Furnace (typically, furnace fan)
- EV charging
- Other home appliances as detected by Sense's disaggregation algorithms.

The CT data comprise, for each CT installed (there could be more than one) 5-minute-resolution data on:

- AC or Furnace runtime (reported in seconds)
- Room temperature (as measured by thermostat, precision varies between homes and thermostat models, typically 0.1°F)
- Thermostat setpoint for cooling and heating
- Ambient temperature (often measured by a dedicated thermometer, i.e., local)

Also, each home's limited metadata (zip code, floor area) was also available. To preserve customer privacy, each home was given an (anonymous) identifier. Figure 1 shows an example of data.

The locations of the selected homes are diverse. Six homes are located in California, two homes in each of Massachusetts and Oregon, and a home in each Arizona, Iowa, Michigan, Missouri, New Jersey, New York, New Mexico, Texas, Virginia and Washington.

Figure 1 shows an example of synergy of the two sources of data (disaggregated from Sense monitor and CT). It can be seen that the power consumption events well correspond to the recorded runtime.

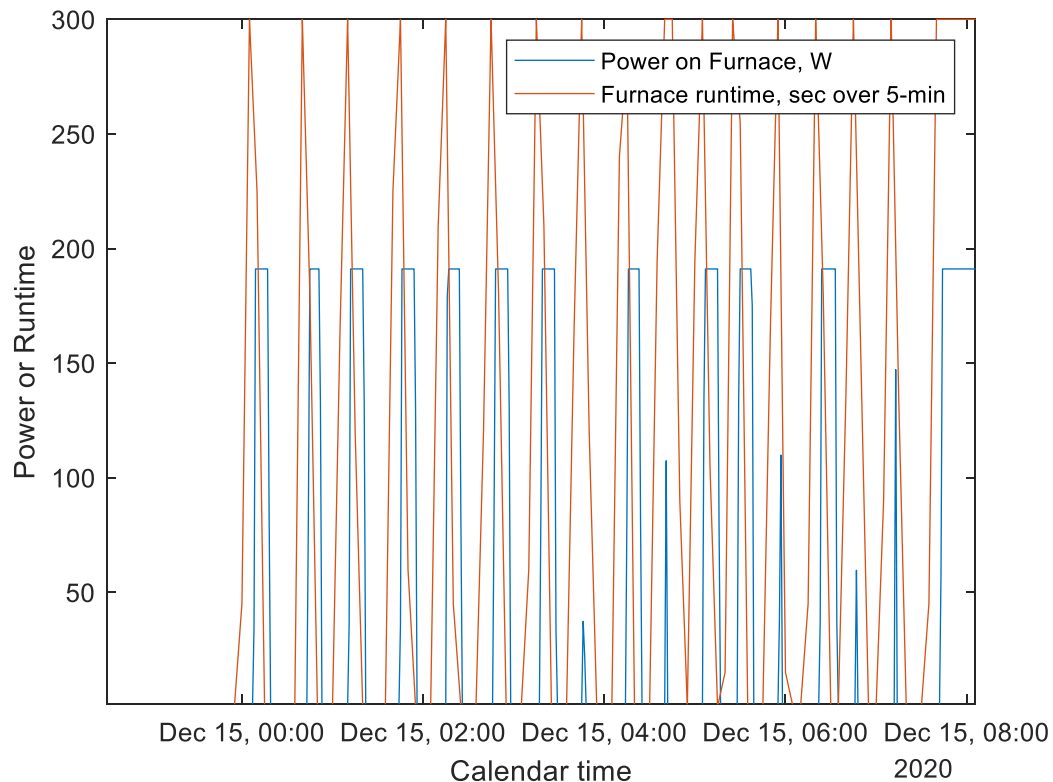


Figure 1. Example of Furnace data obtained by Sense's disaggregating power monitor (power on Furnace's fan) and by communicating thermostat (furnace runtime)

2. Home Thermal Model

Previous Work

Our building thermal model [Zeifman et al, 2019; 2020] for sensible loads in heating season is based on a conventional second-order grey-box model with two capacitances (indoor space and a wall that models entire building shell) given by

$$C_r \frac{dT_r}{dt} = Q_{HVAC} + q_{int} + A_w/(R_w/2)(T_w - T_r) + q_{inf} \quad (1)$$

$$C_w \frac{dT_w}{dt} = A_w/(R_w/2)(T_r - T_w) + A_w/(R_w/2)(T_a - T_w) + q_{ext} \quad (2)$$

where variables T_r , T_w , T_a are, respectively indoor (room), wall and outdoor temperatures, R_w and A_w are overall R-value and building shell area, C_w is overall wall thermal capacitance (thermal mass), C_r is heat capacitance of the internal space, Q_{HVAC} is HVAC heat supply delivered to the conditioned space, q_{int} is internal heat gains/losses (including solar gains through glazing elements) affecting directly T_r , q_{ext} is heat gains/losses affecting the walls externally (i.e., solar gains through opaque elements) and q_{inf} is heat loss due to air infiltration that models both wind and stack effects:

$$q_{inf} = -\rho_{air}c_{p,air}(C_1W^{2.6} + C_2|T_a - T_r|^{1.3})^{0.5}(T_r - T_a) \quad (3)$$

where the first two terms designate air density and heat capacitance, W the wind speed and C_1 and C_2 are fitting parameters.

To improve parameter estimation, we extended the conventional approach by restricting the parameter search space. To this end, we obtain correlations between the runtime fraction and average $(T_r - T_a)$ by integrating Eqs (1)-(2) over a 24-h period of time (unsurprisingly, the correlations are PRISM-like [Fels, 1986], though Eq. (3) predicts some nonlinearity in correlations) and used these correlations to obtain confidence intervals for R-value. In doing so, we assume that the internal and external heat gains are not correlated with the indoor-outdoor temperature difference, thus these quantities integrated over the 24-hours can be considered as random noise. For the air leakage parameters, the key idea is to compare HVAC runtimes for time windows with essentially different wind speeds but similar otherwise (e.g., in terms of time of the day, inside/outside temperatures), in this way we obtain a confidence interval for parameter C_1 . With the parameter search space limited by these confidence intervals, the remaining parameter values can be conventionally estimated, e.g., by curve fitting [Zeifman et al, 2019].

By restricting the parameter search space using correlations obtained over entire heating season, we essentially incorporate all-season data in parameter estimation, making the estimation process robust. This is in contrast with taking an arbitrary two-week period for training as in [Jin et al, 2017], or taking ten most extreme-weather days and repeat model training every month as in [Siemann, 2013].

Runtime Prediction

The main objective of our previous work was estimation and classification of the overall building shell R-value (to determine whether a significant insulation retrofit opportunity exists or not) and estimation and classification of air leakage parameters (air change per hour at 50 PA – ACH₅₀) to identify promising air-sealing opportunities. In this project, we need to characterize

how changing the thermostat setpoints (i.e., the HVAC control) affects the thermal comfort of home occupants (i.e., T_r) and HVAC energy consumption. A conventional noncondensing furnace is probably the easiest HVAC system to model as its heating output is approximately constant at the “on” state and zero at the “off” state. Prediction of T_r and furnace runtime of the selected homes (see section 1) in heating season is main objective of this work.

To make such predictions continuously, we need to quantify the heat gain terms in Equations (1) and (2). The local solar irradiance required for modeling [ISO 13790] of both q_{int} and q_{ext} , is no longer publicly available as part of weather observations and forecasts (see next Section). Fortunately, the local solar irradiance is a prime factor in a home’s PV generation that we have available for each home (see section 1). In this work, we use the first approximation

$$q_{int}(t) = k_1 PV(t) \quad (4)$$

$$q_{ext}(t) = k_2 PV(t) \quad (5)$$

where $PV(t)$ is the time-dependent PV generation of home’s rooftop PV system and k_1, k_2 are fitting parameters. Equations (4) and (5) are related to Eq. (6) that predicts PV generation on an arbitrary oriented surface as a function of solar irradiance and the solar/surface angles¹ (see next Section). In principle, the solar angles can be calculated for a given time and ZIP code, and the PV element orientation can be deduced by calibrating Eq. (6) using the clear-sky periods of time (see next Section). Having these estimates available, we could then calculate the home-level GHI. On the other hand, it can be shown that, for daylight hours with relatively high PV generation and a south-oriented rooftop PV system, Eqs. (4) and (5) are approximately correct. We will consider a more accurate model based on consideration of wall/roof orientations and commercially-available solar irradiance data in future work.

The differential equations (1) and (2) can now be solved numerically and the unknown parameters could be, in principle, estimated by curve fitting [Siemann, 2013; Jin et al, 2017]. That said, the underlying optimization surface could be very shallow, with small changes/uncertainty in the inputs (i.e., interval data) causing large and unpredictable changes in the estimated parameter values, making the parameter estimation problem essentially an ill-posed problem. We believe this ill-posedness of the parameter estimation problem is the main reason for the observed T_r prediction breakdown (see Introduction) and for necessity to regularly re-estimate the parameters [Siemann, 2013].

To alleviate this problem, we build off of our previous work as follows:

- Estimate the building shell surface based on the house floor area.
- Assuming Q_{HVAC} is known in Equations (1)-(2), use the previously developed method (see section Previous Work) to estimate the R_w, C_1 and C_2 . Since these parameters linearly depend on Q_{HVAC} , we will need just to scale them in the subsequent curve fitting.
- Using as much data as numerically possible (30 days), estimate the remaining parameters (C_r, C_w, k_1, k_2) and Q_{HVAC} by fitting the predicted and observed room temperatures.
- For the initial wall temperature, assume the “equilibrium” value (i.e., average between indoor and outdoor temperatures). Given that the “time constant” of homes typically does not exceed a few days [Hedbrant, 2001], this assumption does not affect the parameters estimated with a 30-days prediction horizon.

¹ Note that the solar irradiance on a building element (to which the heat gain is linearly proportional) depends on the element orientation and tilt in the same way as the PV generation depends on PV panel’s orientation and tilt in Eq. (6) [see ISO13790]

- The differential equations are solved using the ode23 solver in MATLAB, and the optimization problem in curve-fitting – by a custom grid-based optimization code.

Our experimentation with optimization in parameter estimation suggests that the proposed gradual estimation process is robust and, unlike the conventional method, yields consistent estimates. As an example, consider a home in California with floorspace of approximately 2,000 ft². Figure 2 shows the correlation between the daily average temperature differences (indoors-outdoors) and the daily furnace runtime fraction. Significant scatter is clearly observable in the Figure. Quantitatively, the correlation coefficient is 0.87, which is lower than the typical correlation coefficients were obtained in the previous work (.92+) and also lower than those calculated for other homes in our dataset in this work. We intentionally selected this home with relatively coarse correlations to check the applicability of the proposed method to a broad range of homes.

The application of the proposed method resulted in the following estimated parameter values: $A_w = 560$, $Q_{HVAC} = 50,000$, $C_r = 50,000,000$, $C_w = 121,600,000$, $C_1 = 0.005$, $C_2 = 0.002$, all in SI units, and dimensionless parameters $k_1 = 5$ and $k_2 = 30$. Figure 3 shows an example of the room temperature prediction (i.e., a solution of the differential equations (1)-(5) with these parameter values) for three days in November 2020. It is seen in the Figure that, although there is a good correspondence between the predicted and observed temperatures, errors of up to 2°C are also seen. Overall, for the entire heating season, the average observed error was 1.05°C, which is higher than the ~0.5°C obtained in [Siemann, 2013]. Noting also the relatively coarse temperature-runtime correlations (Figure 2), we attribute this higher error to additional physical effects, e.g., to the nighttime radiative cooling that could be significant in dry climates, that are neglected in Eqs. (1)-(5). On the other hand, the prediction horizons in both references were small as compared to ours (effectively 24 hours in [Siemann, 2013] and 8 hours in [Jin, 2017]). We intend to incorporate more physical effects in future work.

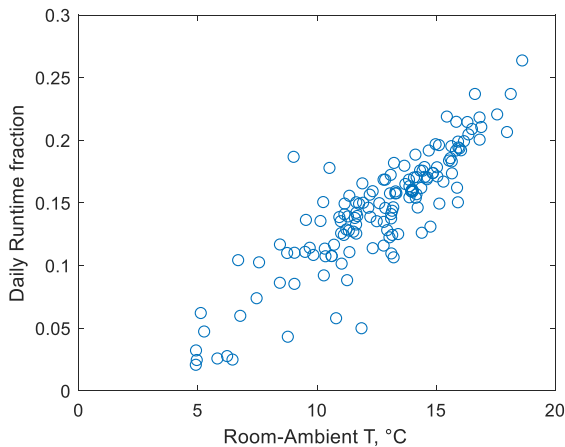


Figure 2. Daily delta temperature - furnace runtime correlations calculated for a home in California using data obtained over winter 2020-2021.

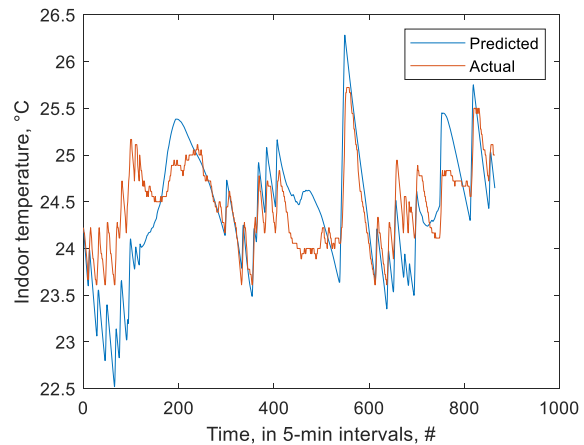


Figure 3. Predicted and actual indoor temperature T_r for the home from Figure 2. Days 19-21 are shown. The overall prediction horizon is 30 days

Prediction of furnace runtime implies simulation of a thermostat that controls the furnace, i.e., the furnace starts whenever T_r falls below the setpoint (minus thermostat's deadband) and stops as soon as the room temperatures grows above the setpoint. The numerical solution of the Equations (1)-(5) with the previously estimated parameters can be used for the simulations. Once the furnace runtimes are calculated using the recorded temperature setpoints and weather data in this way, we can compare, e.g., daily totals of predicted and observed runtimes.

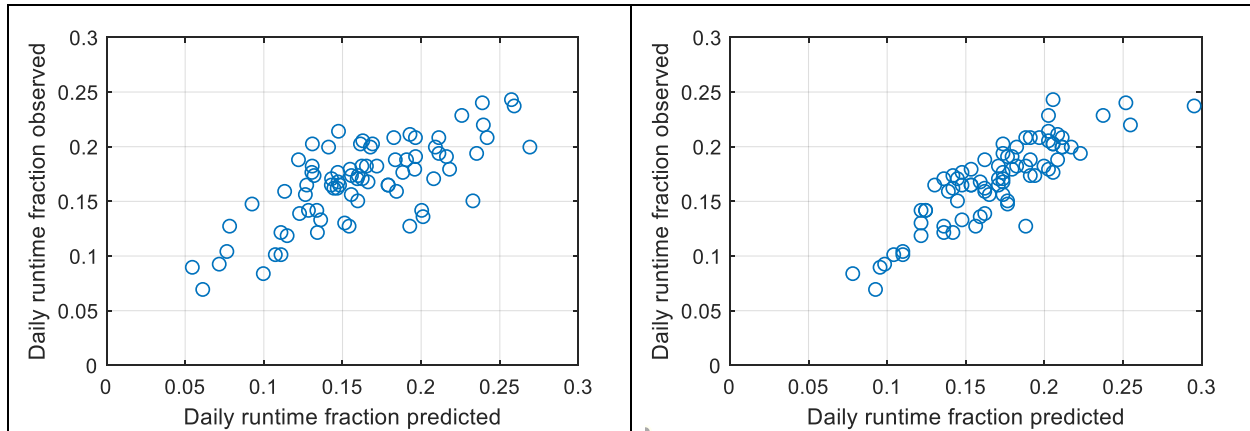


Figure 4. Predicted and actual runtime fraction for the home from Figure 2. Left: previous work. Right: this work.

We could also predict the daily runtime totals using the previous work, i.e., the correlations from Figure 2. Figure 4 shows the results of both methods. It is apparent from the Figure that the predictions based on this work are closer to the experimental values than those based on the previous work. Quantitatively, the average absolute error in predicted daily runtime based on the previous work is 16.9%, whereas that based on this work (estimation of all parameters of the second-order greybox model and using the numerical solution for continuous furnace simulation) is 9.9%, an improvement of about 40%.

3. PV Generation Prediction

In the previous reporting period, we successfully developed PV prediction techniques suitable for the clear sky conditions including shadowing. However, attempts to model PV under cloudy sky conditions were hindered by the coarseness of publicly available cloud cover data (in terms of just few degrees of cloudiness and also in terms of low geographical resolution – up to tens of miles).

In this reporting period, we resolved a data problem with a commercial vendor Solcast and were able to download GHI, solar angles, and weather (ambient temperature, humidity, 10-meter wind speed) that matched the Sense data in terms of time (from November 2020) and geographical locations (ZIP codes) of all our homes.

In line with our overall modeling methodology, we use a hybrid physics-based/ machine-learning approach for PV generation prediction. For a single PV module i , the physics-based prediction is given by [Bashir et al., 2019]:

$$P_{pv,i}(t) = I_{GHI}(t) \times k_i \times \left(1 + c_i \times (T_{base,i} - T_a(t))\right) \times [\cos(90 - \theta(t)) \times \sin(\beta_i) \times \cos(\varphi_i - \alpha(t)) + \sin(90 - \theta(t)) \times \cos(\beta_i)], \quad (6)$$

where I_{GHI} is global horizontal irradiance (GHI), k , c , and T_{base} are site-specific parameters, T_a is local outside air temperature, β is the module tilt, φ is module orientation, θ is the Sun's zenith angle, and α is the Sun's azimuth angle. We used a commercial vendor to obtain hour-resolution weather data that include ZIP-code-based GHI and solar angles (both historical and 24-hour prediction).

For a given home, the parameters k , c , T_{base} , β , and φ can be estimated using the collected PV production data and publicly available weather data, assuming the same orientation of the PV modules². Notably, the independent variables (i.e., predictors) are the solar angles, outside temperature and GHI, but not the calendar time, which facilitates the calibration process.

To calibrate the model, we start from coarse estimation of the geometrical parameters. We initially estimate k as the ratio between the maximum measured PV production and the maximum recorded I_{GHI} value. To find initial estimates of module tilt β and orientation φ , we conduct a coarse grid search, applying Eq. (6), modified to exclude temperature effects, with varying combinations of β and φ and select the combination of values that produced the lowest root mean square error (RMSE) when compared to the measured PV production. These coarse estimates are then refined using a nonlinear numerical optimization method. After k , β , and φ are determined, the parameters related to temperature effects, c and T_{base} , are estimated using a similar process.

We partially validated this technique by applying it to PV production data volunteered by one of the authors, achieving an RMSE of 526 W (maximum PV production for this home is approximately 4500 W). The author in question has confirmed that the parameter estimates obtained by this method for their residential PV modules are relatively accurate. Yet, the overall prediction error of the physics-based approach was relatively large, of the order of 25-30% in terms of the mean absolute percentage error (MAPE).

To further improve prediction accuracy, we employ a machine learning model, a support vector machine regression (SVMR) with a radial basis function kernel. SVMR has been widely used for PV prediction and has been shown to outperform simpler regression-based methods (see, for example, [Alfadda et al. 2017]). In PV applications, the predictors in SVMR are usually time-related variables (e.g., month, hour of day, weekday/weekend) with only few references that use more physically sound solar angles as predictors [Li et al, 2016, Bashir et al, 2019].

In addition to the solar angles, we take the $P_{pv}(t)$ calculated using the calibrated Eq. (6) and GHI values as the predictors in SVMR. Following a cross-validated grid search for hyperparameter tuning, the SVMR is trained and validated on approximately one year's worth of PV production data. We applied this method to the 20 selected homes (see Section 1). With the exception of three homes, the MAPE values, calculated over the six daily peak hours, are 16% or lower for all homes. One of the three homes with high MAPE experienced a twofold increase in PV production midway through the one-year data collection period; another experienced a 50% drop in PV production for approximately two months; and the third experienced intermittent periods of zero PV production as well as a sudden two- to three-fold increase partway through

² In principle, we can consider a mixture model to address multiple module orientations

the year. Thus, we believe that our hybrid physics-based/ SVMR method provides good prediction accuracy for homes with consistent PV production.

Shading

In principle, the shading is addressed by using the solar angles as the SVMR predictors [Bashir, 2019]. However, shading may affect PV generation over only specific times and/or seasons, which leads to under-representation of the shading data in a training data set. To enhance accuracy of shading modeling, we propose a data-driven approach.

Down-selecting the PV production data to clear-sky days, we then employ kriging, a method of spatial interpolation that preserves the 2D auto-correlation structure of data [Cressie, 1990], to model PV production including the times when shading effects are present. The kriging model takes the solar angles α and θ as predictors for the dependent variable P_{PV} . By using the solar angles as inputs, the model can automatically learn when shading occurs (assuming the shading objects are stationary). This is shown in Figure 5 for a sunny day in February 2022, where heavy shading is present in the afternoon.

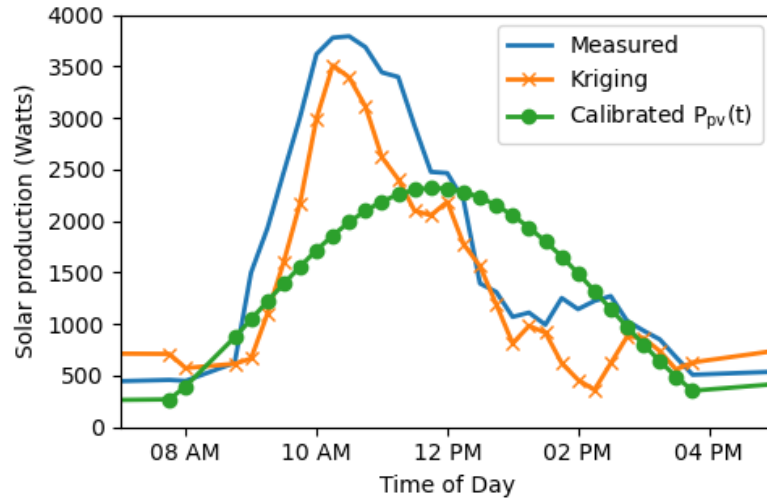


Figure 5. Kriging prediction for a sunny day with heavy shading as compared to the physics-based model given by Eq. (6).

By incorporating the kriging model into the hybrid physics-based/machine learning method explained above (e.g., as an additional predictor variable), we can further increase the prediction accuracy of our PV model. This is ongoing work.

4. Coordination Agent

Using naïve pricing strategies to decrease the peak load in a power-grid may not yield the desired results when the individual participants in the grid services schedule their appliances by solving a similar optimization problem. For example, simultaneous scheduling of flexible appliances towards the time slots with lower electricity prices leads to the undesirable rebound peaks [Li et al, 2012]. To avoid such a scenario, we are proposing an optimization framework analogous to Stackelberg [1934] games.

Consider a game where the players are a coordination agent C , and N individual homes each denoted by H_i . Let p be a vector denoting the price of electricity spanning the next 24 hours, and the coordination agent sets this parameter by minimizing its loss function $\mathcal{L}^C$. Assume

that there are M different appliances at each home, and the power consumption of those appliances must be determined in response to the electricity price p by each home participant. Let $\mathcal{P}_{ij}^*(p)$ denote the power used by appliance j in home i over the horizon K in response to the electricity price p , and each $\mathcal{P}_{ij}^*(p)$ vector forms the j^{th} row of matrix $\mathcal{P}_i^*(p)$. This matrix is computed by minimizing the loss function $\mathcal{L}^{H_i}$ for each home i . Before defining the loss functions explicitly, let us assume that the coordination agent is the leader of the Stackelberg game; the corresponding nested optimization problem takes the following form:

$$\begin{aligned} \min_p \quad & \mathcal{L}^C(p, \mathcal{P}_1^*(p), \mathcal{P}_2^*(p), \dots, \mathcal{P}_N^*(p)) \\ \text{s. t.} \quad & \mathcal{P}_i^*(p) = \underset{\mathcal{P}_i}{\operatorname{argmin}} \quad \mathcal{L}^{H_i}(p, \mathcal{P}_i) \quad \forall i. \end{aligned} \tag{Problem 1}$$

Since the followers set their power schedules as a best response to the price, the scheduled power consumption of each follower depends on the leader's parameters. Since the leader has perfect knowledge on the preferences of houses, the leader can adjust the prices appropriately. Eventually, this optimization structure may help reduce the observed peak loads.

Although the main motivation of the proposed optimization structure is to shape the peak loads, the coordination agent also considers the overall welfare cost of the coordinated homes as well as deviating from the targeted aggregate power consumption. Hence, the loss function of the coordination agent, $\mathcal{L}^C$, consists of two terms. While the targeted aggregate power consumption is represented by a vector $Q = [Q(0) \ Q(1) \ \dots \ Q(K-1)]$, auxiliary variable $u_{ij}(t)$ represents the deviation from the desirable load curve for home i , e.g., $\mathcal{P}_{ij}(t) = \overline{\mathcal{P}_{ij}}(t) + u_{ij}(t)$, where $\overline{\mathcal{P}_{ij}}(t)$ represents the desirable load profile of appliance j for home i at time t . Based on these definitions, the loss function of the coordination agent is given by:

$$\mathcal{L}^C(p, \mathcal{P}_1^*(p), \dots) = \sum_{t=0}^{K-1} \ell_c \left(Q(t) - \sum_{i=1}^N \sum_{j=1}^M \mathcal{P}_{ij}^*(p, t) \right) + \sum_{i=1}^N \sum_{j=1}^M \sum_{t=0}^{K-1} c_j^i \ell(u_{ij}^*(p, t)),$$

where ℓ_c is the agent's loss function, ℓ is a loss function capturing the cost of deviating from the desirable power level, and c_j^i denotes the importance of appliance j for house i .

Similar to the coordination agent's loss function, the loss function of house i consists of two terms. The loss function is structured similar to the "social cost metric" defined in Tsiligkaridis et al. [2019]. However, to ensure that the deviations from the desirable load curve do not violate the individual preferences of the household, the loss function is minimized by taking into account a set of constraints. The resulting optimization problem to be solved by each house i takes the form:

$$\begin{aligned} \mathcal{P}_i^*(p) = \underset{\mathcal{P}_i, u}{\operatorname{argmin}} \quad & \sum_{t=0}^{K-1} \sum_{j=1}^M c_j^i \ell(u_{ij}(t)) + \left(\sum_{t=0}^{K-1} \sum_{j=1}^M \mathcal{P}_{ij}(t) p(t) - \sum_{t=0}^{K-1} \mathcal{P}_{i,PV}(t) p(t) \right) \\ \text{s. t.} \quad & \mathcal{P}_{ij}(t) = \overline{\mathcal{P}_{ij}}(t) + u_{ij}(t) \quad \forall i, j, t, \\ & \mathcal{P}_{ij}(t) \geq 0 \quad \forall i, j, t, \\ & (+\text{Appliances/PV-related Constraints}) \end{aligned}$$

Given that loss functions, $\ell(\cdot)$ and $\ell_c(\cdot)$ are simple absolute value functions, it is possible to formulate a tractable optimization problem. To this end, we add the optimality conditions of individual households as constraints to the coordination agent Problem 1.

Problem 1 can be formulated properly after estimating *the* c_j^i values for each appliance for each home, and past disaggregated load profiles of the participating homes will be utilized to estimate those values [Tsiligkaridis et al, 2019]. In addition to past power consumption data, estimates of the desirable load profiles for the interested horizon and the basic properties of the appliances are required to define the comfort related constraints. The electricity price vector p minimizing the objective of the coordination agent can be determined by optimizing the reformulated problem.

To test the performance of the proposed optimization framework, we designed a simulation environment consisting of 50 homes in Python, using the Gurobi 9.5.1 package [Ozcan and Paschalidis, 2022]. We assume that all considered appliances are available in each home and those appliances are identical across the community, whereas the preferences of households (e.g., thermal comfort interval) are randomly initialized for each house. Finally, the desirable load curve, $\overline{\mathcal{P}}_{ij}$, is randomly generated for each house and appliance, and we obtain the desirable load curve of community by summing $\overline{\mathcal{P}}_{ij}$ curves over each house and appliance. We set the length of time intervals equal to 15 minutes ($K = 96$).

At the current stage, we set the cost of deviating from the desirable load profile for each house with the help of a parameter, b_i , which indicates the maximum amount of money household i is willing to pay for the next 24 hours. If the average cost of consuming electricity as much as the desirable load profiles is higher than b_i , then c_j^i values are initialized in a way to incentivize consumers to deviate from their desirable load curves.

As a first simulation example, we consider naïve optimization (i.e., no social welfare of the community is considered). To this end, we initialize a community and generate a random price vector p . Each household i then calculates its own optimal load curve $\mathcal{P}_i^*(p)$ with respect to p by minimizing $\mathcal{L}^{H_i}$. The optimal power curve of the community, $\mathcal{P}^*(p)$, can be obtained by summing the optimal load schedules of individual households.

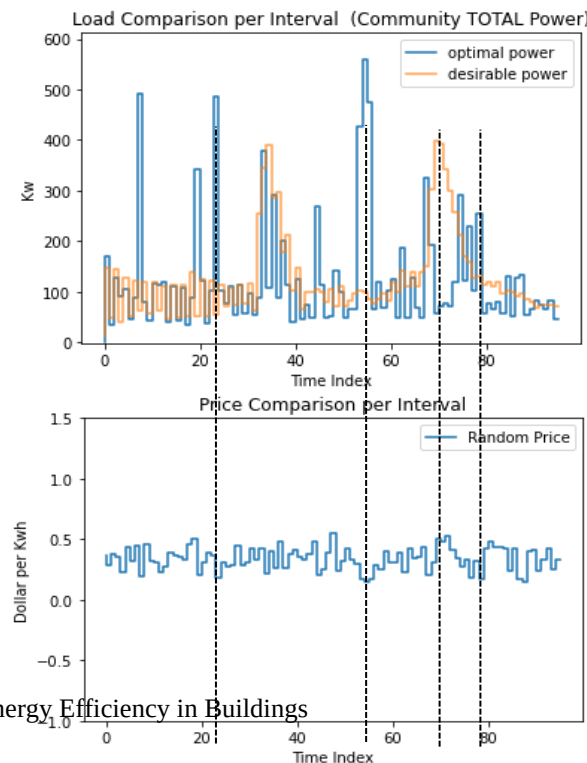


Figure 6. The response of community to a random price vector. Top: Deviation from the desirable power of community (orange line) of the resultant optimal power schedule of community (blue line) with respect to the announced random price vector. Bottom: The random price vector.

Figure 6 shows the calculated response. It can be seen that increasing the price slightly to avoid the peak in the desirable power curve of the community (the peak in the orange line around the 70th time period) causes many more rebound peaks at the locations where the electricity price is low. Hence, it is crucially important to have a coordination agent while setting the prices.

The missing part preventing us from formulating the coordination agent's problem is the targeted aggregate electricity consumption vector, Q , which must be determined by the coordination agent prior to formulating Problem 1. For the next example, we use the same initialized community, but randomly sample the Q vectors from different multivariate Gaussian distributions and optimized the *coordination agent's* problem. Figure 7 shows the change in the optimal power schedule of the community with respect to the different targeted aggregate electricity consumption vectors decided by the coordination agent.

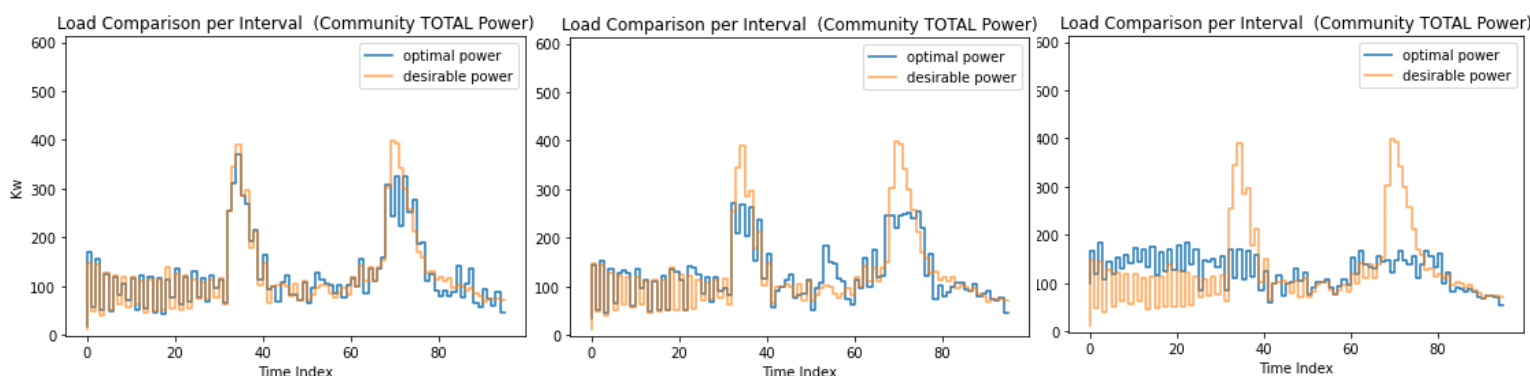


Figure 7. The desirable ($\bar{P}$) and the scheduled optimal real power ($\mathcal{P}^*$) of the community under different targeted aggregate power consumption (Q) vectors in Kw. Left: $Q \sim N(300, \Sigma)$. Middle: $Q \sim N(200, \Sigma)$. Right: $Q \sim N(100, \Sigma)$, Σ is a diagonal covariance matrix with all entries on the main diagonal equal to 10.

From Figure 7, it can be observed that the peaks in the desirable power profile of the community reduce when the coordination agent lowers its targeted aggregate power consumption level. However, the main instrument the coordination agent utilizes while achieving this is the electricity price. In all cases shown in Figure 7, the coordination agent sets its price vector differently in order to minimize its own objective function by considering the optimality conditions of the individual households. Figure 8 shows the change in the electricity prices set by the coordination agent under three different scenarios introduced above.

Figure 8 suggests that the price set by the coordination agent tends to increase when the mean of the targeted aggregate power consumption vector decreases. If we look at this relation from the perspective of classical economics, it sounds natural to consider the targeted aggregate power consumption vector as the amount of the energy that coordination agent wants to supply at each time point to the community. Therefore, we can basically say that decreasing power supply leads into an increase in the price. Apart from this interpretation, we can intuitively guess the change in price vectors by looking at Figure 7. We observe larger deviations from the desirable load profile when the mean of targeted aggregate power consumption vector decreases in Figure 7. Of course, the coordination agent enforces those deviations by increasing the electricity price. Therefore, it is reasonable to expect to see higher prices when the mean of targeted aggregate power consumption increases.

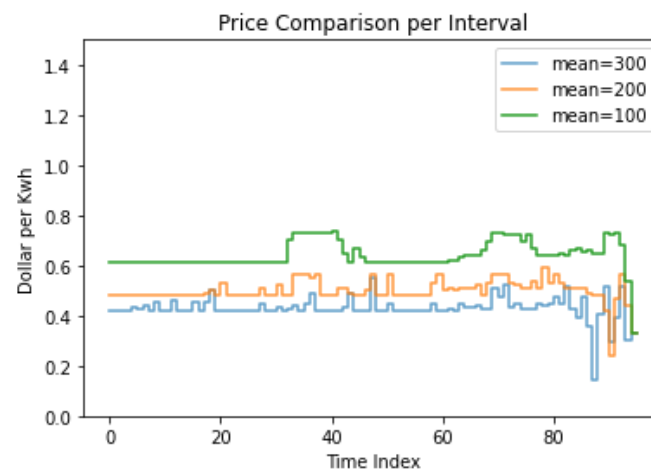


Figure 8. The change of the electricity price with respect to the mean of targeted aggregate power consumption vector Q .

Summary and Future Work

We presented an approach to enable conversion of existing homes into grid-integrated buildings. The homes need to be equipped with nonintrusive ubiquitous sensors, the self-installed Sense energy monitors that disaggregate overall home electricity consumption into loads on individual appliances as well as communicating thermostats. The sensor data streams are used to automatically predict major household electrical loads and their flexibility, which information is fed into a computationally efficient coordination agent that optimally schedules the flexible loads while preserving the occupant comfort.

In this work, we considered about twenty disparate homes equipped with such sensors as well as a rooftop PV system, but without battery energy storage. We extended our approach, developed for quantification of home retrofit opportunities, to model room temperature and HVAC runtime in heating season. In lieu of the generally unavailable solar irradiation data, we use home PV generation data, measured by the energy monitor. We also proposed a robust two-step approach to automatically calibrate the model, i.e., to estimate its parameters. While the approach yields relatively accurate results in terms of runtime error (about 7% on average), it could be improved to yield better prediction accuracy for room temperature. In particular, we

would incorporate nighttime radiative cooling and/or consider a third-order greybox model to better represent heat transfer between the building shell, home hardware content and the air.

Finally, we formulate an inverse optimization method that converts several services (energy efficiency, load shifting, load shading and load shaping) into corresponding objective functions and has linear complexity over the number of homes and over the number of time steps.

In future work, we will extend the proposed approach with models of an additional thermostatic load (DHW) and a user-controlled appliance (EV charger) and incorporate all the models in the coordination agent. The data collected from the considered homes enable development of both modeling and optimization approaches.

Acknowledgements

The work presented in this paper was funded in part by the U.S. Department of Energy Building Technologies Office Building America Program under Contract #DE-EE 0009696.

References

- Alfadda, A., et al, 2017. "Hour-ahead solar PV power forecasting using SVR based approach," IEEE Innovative Smart Grid Technologies Conference (ISGT), pp. 1-5.
- Bashir, N. et al. 2019. "Solar-TK: A Data-Driven Toolkit for Solar PV Performance Modeling and Forecasting," IEEE 16th MASS conference, pp. 456-466.
- Battery Storage. 2022. "Home Battery Storage Explained," <https://www.cleanenergyreviews.info/battery-storage>
- Cressie, N. (1990) "The origins of kriging," Mathematical Geology 22, pp. 239–252.
- 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>
- Fels, M. 1986. "PRISM: An introduction." Energy and Buildings 9, pp. 5–18.
- ISO 1370. 2008. "Energy performance of buildings — Calculation of energy use for space heating and cooling," International standard.
- Johan Hedbrant, 2001. "On the thermal inertia and time constant of single-family houses" M.Sc. Thesis, Linköping University.
- 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, Z. et al, 2016. “A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting.” *Energies* 9, p. 55.
- Li Y., B.L. Ng, M. Trayer. 2012. “Automated Residential Demand Response: Algorithmic Implications of Pricing Models,” *IEEE Transactions on Smart Grid* 3, pp. 1712-1721.
- Lin Y., T. Middelkoop, P. Barooah. 2012. “Issues in identification of control-oriented thermal models of zones in multi-zone buildings,” Figures 3 and 7, *IEEE Conference on Decision and Control*.
- Ozcan, E.C., Paschalidis, I. C. 2022. “energy_Almost_DIY,” source code available at https://github.com/erhancanozcan/energy_Almost_DIY
- Roth, A. 2019. “Grid-interactive Efficient Buildings Technical Report Series,” DOE, <https://www1.eere.energy.gov/buildings/pdfs/75478.pdf>
- Sense. 2022. Sense electric monitor, <https://sense.com>
- Siemann, M.J. 2013. “Performance and applications of residential building energy grey-box models,” University of Maryland. Ph.D. Thesis.
- Von Stackelberg, H. 1934. “Market structure and equilibrium.”
- Sturzenegger D. et al. 2012. “Semi-Automated Modular Modeling of Buildings for Model Predictive Control,” *BuildSys Conference Proceedings*, pp. 99-106.
- Tsiligkaridis, A., Paschalidis, I. C., & Coskun, A. 2019. “Data center demand response pricing using inverse optimization,” In *Proceedings of the Tenth ACM International Conference on Future Energy Systems* (pp. 400-402).
- Zeifman, M., A. Lazrak, K. Roth. 2019. “Residential retrofits at scale: opportunity identification, saving estimation, and personalized messaging based on communicating thermostat data,” *Energy Efficiency* 13, pp. 393-405.
- Zeifman, M., A. Lazrak, K. Roth. 2020. “When Two Isn’t Better Than One: Algorithms to Remotely Identify Insulation and Air Sealing Retrofit Opportunities in Homes with Two Communicating Thermostats,” *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings*, 12-467 - 12-478