

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Artificial Intelligence Method for the Forecast and Separation of Total and HVAC Loads with Application to Energy Management of Smart and NZE Homes

ROSEMARY ALDEN, (Student Member, IEEE), HUANGJIE GONG, (Student Member, IEEE), EVAN S. JONES, (Student Member, IEEE), CRISTINEL ABABEI¹, (Senior Member, IEEE), and DAN M. IONEL, (Fellow, IEEE)

SPARK Laboratory, ECE Department, University of Kentucky, Lexington, KY, USA

¹ ECE Department, Marquette University, Milwaukee, Wisconsin, USA

Corresponding author: Cristinel Ababei (e-mail: cristinel.ababei@marquette.edu).

ABSTRACT Separating the HVAC energy use from the total residential load can be used to improve energy usage monitoring and to enhance the house energy management systems (HEMS) for existing houses that do not have dedicated HVAC circuits. In this paper, a novel method is proposed to separate the HVAC dominant load component from the house load. The proposed method utilizes deep learning techniques and the physical relationship between HVAC energy use and weather. It employs novel long short-term memory (LSTM) encoder-decoder machine learning (ML) models, which are developed based on future weather data input in place of weather forecasts. In addition to being used in the proposed HVAC separation method, the LSTM models are employed also for accurate day-ahead HVAC and solar photovoltaic (PV) energy forecasts. To test and validate the proposed method, the SHINES dataset, a publicly available dataset spanning three years at 15-minute time resolution from a large-scale DOE experimental project, is used. Two computational case studies are constructed with a test HEMS to investigate the power regulating capability of smart home virtual operation as dispatchable loads or generators. Prediction results obtained with the proposed method show hourly and daily CV(RMSE) of 29.4% and 11.1%, respectively. These results are well within the bounds of error established by academia and the ASHRAE building model and calibration standards.

INDEX TERMS Machine learning (ML), Long short-term memory (LSTM), Home energy management system (HEMS), Demand Response (DR), Solar photovoltaic (PV), Non-intrusive load monitoring (NILM), Heating, ventilation and air-conditioning (HVAC) systems, Distributed energy resources (DER), Smart home, Smart grid

I. INTRODUCTION

RESIDENTIAL demand of modern state of the art buildings is a large and growing portion of total electricity usage. According to the U.S. Energy Information Administration (EIA), population increase, urbanization, access to electricity for end use purposes such as appliances and space conditioning, and rising use of electric vehicles are factors driving this trend. With this increase in residential energy usage, innovative planning with home energy management systems (HEMS) is necessary to reduce the increasing carbon footprint by incorporating distributed energy resources

(DERs) such as solar photovoltaic (PV) and battery energy storage (BES) systems. Research explores demand response (DR) schemes to incorporate solar PV, BES, and HVAC system controls into modern HEMS for the smart grid. Many of these studies involve machine learning (ML) based forecasting employed in look-ahead optimizations. Previous studies may be classified into two main categories: community/district and residential level. Community-level studies are more prominent and have reported electricity savings through smart home control methods such as HVAC system load scheduling.

Examples of community-level studies of HVAC load control include the work by Gong *et al.*, which presents an aggregation technique used for reducing power peaks in a typical community constructed with reference homes [1]. As another example, Bandyopadhyay *et al.* presented a K-means clustering algorithm to shift the peak loads of the HVAC system, electric water heater (EWH), and pool pumps outside of high demand hours [2]. The study completed by Oakridge National Lab in [3] presented an investigation into adaptive controls and Model Free Controls (MFC) to control a community of HVAC systems load demand to follow a daily profile of PV generation.

Residence-level studies into HVAC controls are much less common in part due to the challenges of energy forecasting from increased variability or unpredictability when compared to aggregated load forecasting at the community-level. An example of such studies is the work of Kurte *et al.* that presented a deep reinforcement learning based HVAC control approach for residences which reported 30% cost reduction through simulations and 21% reduction on real houses [4].

Often times, the above studies make use of various forecasting or prediction techniques to implement look-ahead - such as day-ahead - type of optimizations. One can find such forecasting techniques for PV generation, HVAC, or total house or community power loads. For example, PV generation forecasting was studied with both machine learning and physics based models. Hafiz *et al.* described an energy management system to optimize energy purchasing cost and to reduce peak power usage for residences based on solar-PV forecasts using a Long Short-Term Memory (LSTM) model [5]. The study in [6] compared several forecasting methods for PV generation and reported that LSTM models provided the lowest prediction errors.

Many previous studies focused on HVAC load forecasting for combined industrial or commercial buildings, which exhibit smoother load curves, and therefore are simpler to predict. An example study in this more mature area include the Artificial Neural Network (ANN) activation layer comparison and sensitivity study of weather inputs in the work by Zwolinska *et al.* on a hotel building [7]. Lu *et al.* implements an LSTM encoder-decoder layer configuration method with an additional attention mechanism layer to predict a regional community conditioning load including weather inputs [8]. The investigation conducted by Klessi *et al.* [9] also reported that LSTM models outperform ANN models in forecasting energy usage. Their paper compared the standard LSTM model and the more advanced encoder-decoder configuration to showcase the even better performance of the advanced configuration on data sets from an amusement park building, an office building, and an operation center building. Kim *et al.* [10] also used the LSTM encoder-decoder model to predict simulated heating and cooling load from an EnergyPlus model of an office building.

Studies have found that LSTM models were among the best performing at residential load prediction. Turgut *et al.* showed that LSTM models outperformed other models

when forecasting appliance power such as the television and microwave [11]. The study in [12] also investigated LSTM models to predict total load and lighting for several different prediction windows or horizons. In another recent study, Syed *et al.* investigated a bi-directional LSTM model and compared it to other configurations with the use of future weather inputs to predict the household and appliance power at a high resolution of 10 minutes [13]. This reference does not forecast the air-conditioning load on its own or disaggregate it from the total load, which is a main focus of this current paper.

On the other hand, one can find only a few studies on home-level air conditioning load prediction, and, those too identified LSTM recurrent neural networks (RNN) as the most promising ML approach. For example, Gutierrez *et al.* used an LSTM encoder-decoder model to predict the daily HVAC consumption of a smart home [14]. Xia *et al.* also used a bi-directional LSTM model for daily HVAC predictions [15]. Zingre *et al.* used LSTM models in conjunction with EnergyPlus output data to predict cooling load at a finer resolution [16]. These studies reported a variety of prediction error metrics, which makes their overall comparison challenging and emphasizes the need for further standardization of error reporting in this context. Along this line, the authors also propose in this paper to adopt the Coefficient of Variation (CV) of the Root Mean Square Error (RMSE) error metric to align with the ASHRAE Building modeling standards, denoted throughout the paper as CV(RMSE).

Recent studies integrated LSTM models with other methods such as Sparrow Search Algorithm (SSA) swarm intelligence optimization [17] or Particle Swarm Optimization [18]. Motivated by these previous studies, in this paper, an investigation into novel LSTM models for prediction of PV generation and HVAC load for an individual residence is completed, and those models are used for HEMS HVAC load dis-aggregation computational case studies. More specifically, to gain insights into the benefits of machine learning models for residential HVAC power separation, this paper's objectives are to provide contributions to the following areas: (1) development of satisfactory LSTM models for HVAC load with future weather inputs for prediction of single-residence HVAC and PV system power; (2) selection of physical characteristics for PV forecasting based on typical PV power equations and differential evolution (DE); (3) novel separation method of HVAC power from total residential load; (4) combined HEMS control method that incorporates HVAC, PV, and BES systems by employing the LSTM forecasting.

The remainder of this paper starts with a discussion on the experimental data set used for developing the proposed encoder-decoder LSTM model for HVAC load prediction, which is provided in Section II. Then, the LSTM model is presented in Section III, and the classic physics-based method for predicting the PV generation is discussed in Section IV. Sections V and VI provide the prediction results and the proposed novel HVAC load separation method, respec-

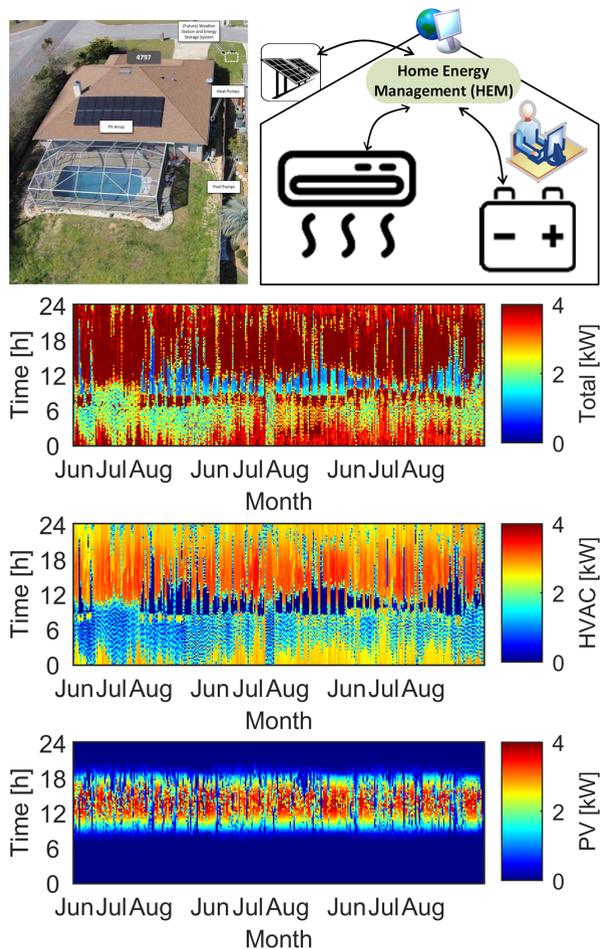


FIGURE 1. SHINES House #1 [19] and schematic of the components and HEMS considered in the study (top). Experimental data available for total and HVAC load and PV generation over three years 2018, 2019, and 2020 (bottom). Seasonal summer sets for daily and day-time (6am to 9pm) data have been employed in the reported studies.

tively. The behind-the-meter (BTM) optimal control studies that utilize the LSTM forecasts are included in Section VII, and conclusions are discussed in the final section.

II. THE EXPERIMENTAL DATA

This paper utilizes experimental data from the SHINES Residential Demonstration, a field site comprised of two homes in Pensacola, Florida managed by the Electric Power Research Institute (EPRI). Each home is equipped with rooftop solar PV and BES systems. Average electric power data with a resolution of 15 minutes is available for the local distribution transformer and total house usages as well as for the water heaters, pool pumps, and HVAC, solar PV, and BES systems through the data dashboard [19]. Local weather data, including ambient temperature and solar irradiance, are also provided at the same resolution.

The SHINES power and weather data for the summer months, specifically June 1st to September 15th, was selected for years 2018-2020 from House #1 as illustrated in Fig 1. In addition to the full 24-hour data set, a “pruned” version of

the summer data was processed such that only times from 6am to 9pm were considered. This method of “pruning” isolates the time of particular interest during which HVAC system demand significantly increase total energy load to be met by utilities. This time frame also includes the times at which solar PV generation is commonly available. The combination of high HVAC system demand and solar PV generation offers an opportunity for HEMS control to be utilized most effectively.

Section V shows that the discussed energy forecasting methods produce satisfactory results when utilizing the “pruned” dataset. In the following section, the “non-pruned” data set with full 24-hour days is employed to show the effectiveness of the novel separation method for HVAC system power. The resulting advantage of no longer requiring dedicated instrumentation to monitor HVAC system energy is also discussed. It should be noted that, for both data sets, a conversion from 15-minute to hourly resolution was performed to reduce computational runtime and to reduce variability of ML inputs while retaining both the trends and patterns captured by higher resolution data.

III. ENCODER-DECODER LSTM MODELS FOR SEQUENCE-TO-SEQUENCE PREDICTION OF HVAC LOAD AND PV GENERATION

LSTM models are a category of ML models within RNNs. Generally, RNNs are different from standard neural networks (NNs) in that they have additional connections or loops in the topology between cells in a layer. It is these connections that enable RNN models to remember past information by adding memory or state to the network. The ability to accumulate state over the input sequence allows the RNN model to learn the ordered nature within input sequences’ observations. However, standard RNN models may suffer from the so-called vanishing gradient problem that can degrade performance.

It was reported that the encoder-decoder approach to sequence prediction is more effective compared to outputting a vector directly [20]. Therefore, an LSTM model for multi-step sequence-to-sequence forecasting with an encoder-decoder two-layer structure based on [21] was developed and the architecture is shown (Fig. 2). The model is used to predict both residential HVAC system load and PV generation.

The encoder sub-model receives the input sequence and produces a compressed fixed-length vector, as an internal representation of the input. A stacked LSTM model is constructed by using multiple hidden layers stacked together. The encoder’s output is repeated a number of times equal to the number of steps for which the output sequence must be predicted and then used as input into the decoder sub-model.

The decoder sub-model generates an output value for each predicted output time step. The model from Fig. 2 predicts a sequence of output time steps with the entire prediction being produced at once as a vector of k values, which will be interpreted as time steps. In other words, there are several parallel input time series (i.e., multiple observations at the

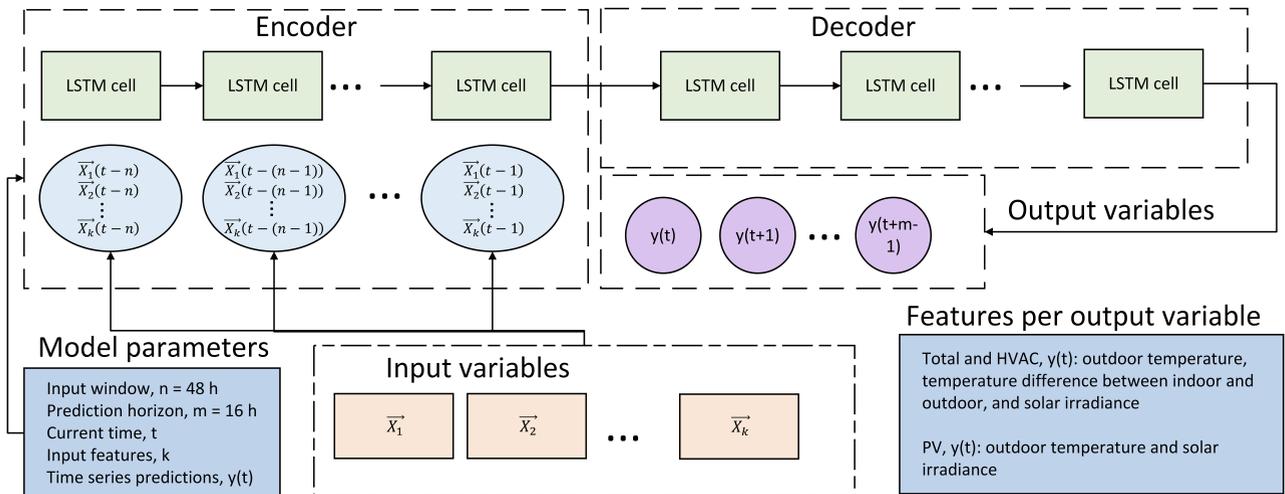


FIGURE 2. General diagram of the architecture for the proposed LSTM model. Actual weather data was used as an input to emulate a "perfect" weather forecast so that the best possible performance may be determined by utilizing the relationships between the weather, HVAC energy usage, and PV generation at the time of predictions.

same time step) and an output time series, which depends on the inputs. This is a forecasting problem, which is formulated or framed as a *multi-step multivariate input, multi-step univariate output* time series forecasting problem.

The LSTM model was trained with data from the first two summers of the SHINES experimental dataset, and the predictions were tested on the data from the third summer of the dataset. The prediction horizon or output for the residential HVAC power load and PV generation was the next 16 hour "pruned" day, from 6am to 9pm. The model was provided with two days of previous power usage/generation as well as two days previous and one day future of selected weather conditions. The one day future input plays the role of a weather forecast to capture the impact of weather on HVAC and PV power. Therefore, the weather input is a "perfect forecast" as actual measured data is used in its place for validation of the approach.

The types of model-inputs used in previous air conditioning studies highlighted the importance of solar irradiance and temperature in capturing the HVAC usage patterns [22]. In addition, the study by Liang and Ma [23] explained the influence of thermal inertia in a home and the concept of a "lag" between the outdoor temperature and the peak HVAC usage. Informed by results from previous studies on ML based forecasting and by the physical relationships between weather conditions and residential HVAC load, in this encoder-decoder LSTM model, the outdoor temperature, temperature difference from an internal set-point (assumed to be 21 °C for this home) to model the thermal inertia of the home, and solar irradiance were selected as inputs into the model for HVAC load predictions. For the PV generation predictions, the inputs include the outdoor temperature and solar irradiance, which were found to be among the most relevant by previous studies as well [5].

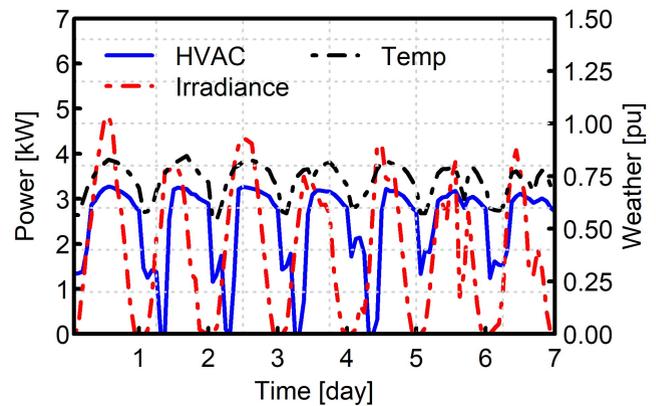


FIGURE 3. The thermal inertia of this home can be seen in this example week of August 2nd to August 8th. The HVAC system load starts to increase after the temperature and irradiance start to increase, delayed about one hour.

IV. PHYSICS-BASED CONSIDERATIONS AND PV GENERATION CALCULATOR

The AI models presented in this paper include inputs that are based on the physical relationships between the predicted generation or load to the weather, ie outdoor temperature and solar irradiance, at the time of the prediction. Another important physical relationship that the models must capture for suitable HVAC load predictions is the thermal inertia of the home, ie the capacity or ability to store heat or insulate from external heat, which can be seen in the delay between when the outdoor temperature starts to increase and when the HVAC systems starts drawing power to cool the home (Fig. 3).

The machine learning model learns this thermal inertia based relationship to the outdoor temperature and solar irradiance to have accurate prediction results for the cooling load. The PV predictions are not influenced by a lag-time

effect from weather inputs and may thus be easier to forecast using a machine learning or calculation method.

To provide a basis or reference of comparison for the accuracy of the proposed ML based prediction models, a classic physics-based method for predicting PV generation was developed. The power estimator utilizes the onsite, measured outdoor temperature and irradiance as inputs and was developed for the rooftop solar PV system of the EPRI SHINES House #1. The power output of the PV power estimator and the net power flow are estimated by the following expressions:

$$P_{ac} = \left[\left(\frac{\gamma \eta}{1000} \right) P_r \right] \left[1 - \left(\frac{k_p}{100} (T_{cell} - 25^\circ C) \right) \right] \quad (1)$$

$$T_{cell} = T_{amb} + \frac{NOCT - 20^\circ C}{0.8} \left(\frac{\gamma}{1000} \right) \quad (2)$$

$$P_N = HVAC - PV + M + B \quad (3)$$

where P_{ac} is the AC power output [W], γ , the solar irradiance [W/m^2], P_r , the rated PV array DC power [W], k_p , the temperature coefficient of maximum power [$\%/^\circ C$], T_{cell} , the temperature of the PV cell [$^\circ C$], T_{amb} , the outdoor ambient temperature [$^\circ C$], NOCT, the nominal operating cell temperature [$^\circ C$], and η , the efficiency of the system. The net total power, battery power, and miscellaneous loads of the home are denoted by P_N , B, and M for the demand response case study. The accuracy of the estimator was improved through a DE based optimization scheme with the objective of determining the best feasible combination of parameters P_r , k_p , NOCT, and η based on the Root Mean Square Error (RMSE) between the estimator output and measured PV power data.

V. PREDICTION RESULTS AND ERROR ANALYSIS

This section presents prediction results obtained with the LSTM model introduced in the previous section. These results include forecasts of the HVAC load and PV generation during the summer of 2020. Representative plots of these forecasts are shown in Fig. 4 and Fig. 5, where it can be seen that the forecast traces follow closely with the actual measurements.

The plot from Fig. 5 also shows the PV generation as calculated by the classical model paired with DE optimization discussed in Section IV. It can be noted that the LSTM based forecasting performs comparably to the estimation method, used as a benchmark here, indicating that machine learning methods such as the LSTM can well serve as a day-ahead alternative to classical real time estimations.

To quantify the forecasting accuracy of the LSTM encoder-decoder model for PV and HVAC power, Table 1 reports several common error metrics, which demonstrate the results are satisfactory. One of the metrics included, the covariance (CV) of the Root Mean Square Error (RMSE), is a common building modeling error calculation that can be

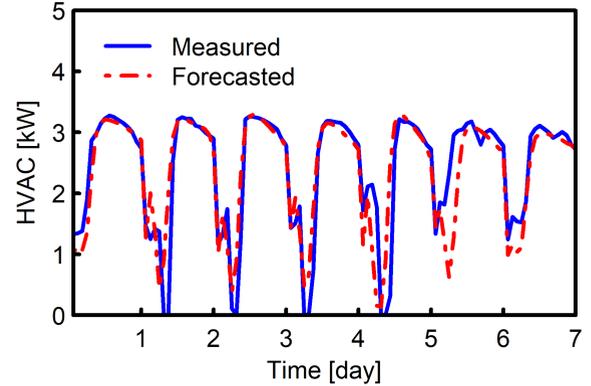


FIGURE 4. Day-ahead, hourly forecast of HVAC power draw during the example week also showed in Fig. 3. Daily and hourly CV(RMSE) errors are substantially lower than ASHRAE requirements for a calibrated building energy model.

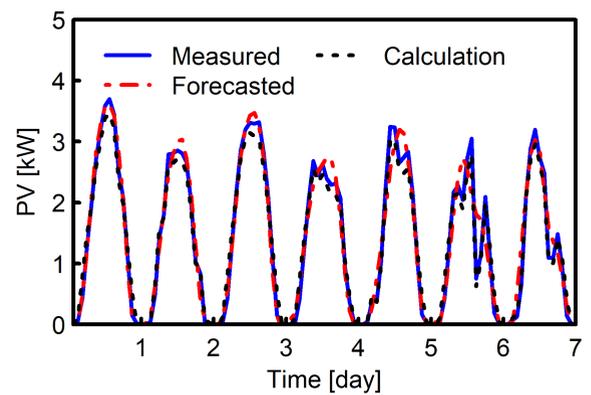


FIGURE 5. Day-ahead hourly PV generation forecast from the same week in August for uniformity. The PV day-ahead forecasts are comparable with the calculation for that timestep, indicating a very strong performance of the LSTM model against the benchmark.

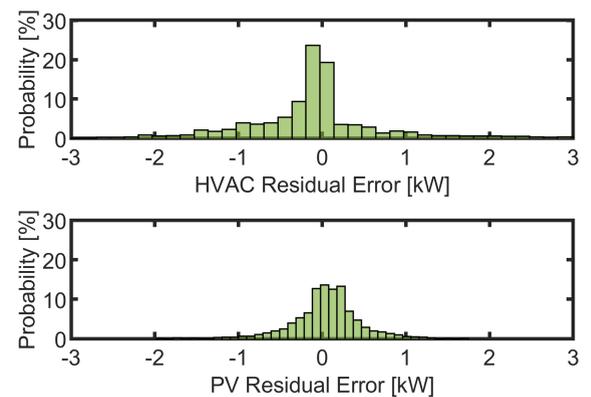


FIGURE 6. The errors for the HVAC forecasts over the entire test period of three and a half months are strongly clustered with approx. 45% probability within ± 0.2 kW around 0. The PV forecasts are also strongly clustered with the majority of the error within ± 0.25 kW of 0. Results are satisfactory considering the peak power of approx. 3kW for both the HVAC and the PV.

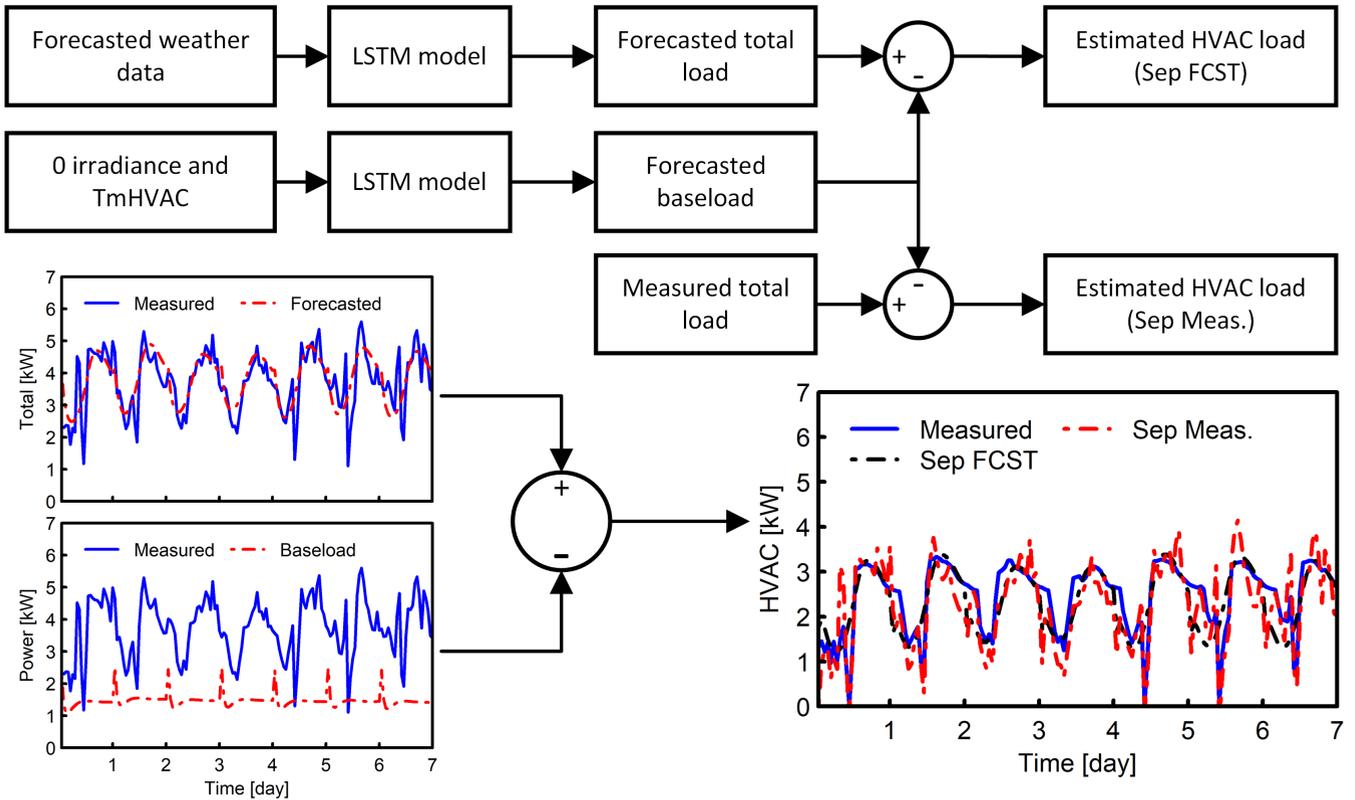


FIGURE 7. Schematic diagram of the mathematical algorithm for the novel HVAC separation method for individual residences. Total load and baseload forecasts are illustrated in the two smaller graphs and an example week from July 16 to July 22, 2020 of the HVAC separation from the total forecasts (Sep FCST) and the total measured data (Sep Meas.) is shown in the larger plot. The separated HVAC load for the summer of 2020 is satisfactory but not as accurate as the forecasts from HVAC measured data shown in Section V.

applied to machine learning forecasting to assess suitability of these artificial intelligence (AI) based modeling methods against other types of building models such as black box or physics based models.

The ASHRAE Standards indicate that at the hourly resolution a CV(RMSE) of 30% is considered calibrated for a building [24] and a consensus between researchers applying the ASHRAE standards to HVAC modeling specifically has determined that a daily CV(RMSE) under 35% is acceptable [25]. The HVAC load predictions obtained with the proposed LSTM model using the “perfect” weather forecast are well within these limits. These predictions will be used later in Section VII.

Furthermore, the HVAC and PV error distributions are very strongly clustered around 0 kW as shown in Fig. 6. For the HVC predictions, a nearly 45% likelihood that an error of 0.2 kW or less occurs. The PV is also clustered near zero, with more than the majority of the errors within 0.25 kW.

VI. NOVEL HVAC LOAD SEPARATION METHOD USING LSTM MODEL

The approach presented in the previous section for HVAC load forecasts was developed under the assumption that the subject modern house is equipped with dedicated HVAC energy monitoring circuits, which collect real HVAC mea-

TABLE 1. Summary of errors for solar PV calculations as well as PV forecasts using the LSTM encoder-decoder method.

Hourly Metric	PV Calc	PV	HVAC
Average Load [kW]	1.186	1.186	2.199
RMSE [kW]	0.149	0.314	0.701
CV(RMSE) [%]	14.89	26.5	31.8
R^2	0.975	0.923	0.486
Daily Metric	PV Calc	PV	HVAC
Average Load [kWh/day]	18.976	18.976	35.197
RMSE [kWh/day]	1.256	1.353	6.624
CV(RMSE) [%]	6.7	7.1	18.81
R^2	0.958	0.947	0.307

surements. These measurements provide both initial training data as well as future data for model refinements.

Such HVAC load forecasting were used in the BTM DR schemes to be discussed in the next section, but also, generally, another application of this energy usage information could be to educate the home owner for behavior change purposes such as to reduce usage and cost. In addition, such information can also be beneficial for grid planning, load estimation, and analysis purposes to utilities. However, many homes constructed or in use today are not equipped with dedicated HVAC energy monitoring systems and, thus, the

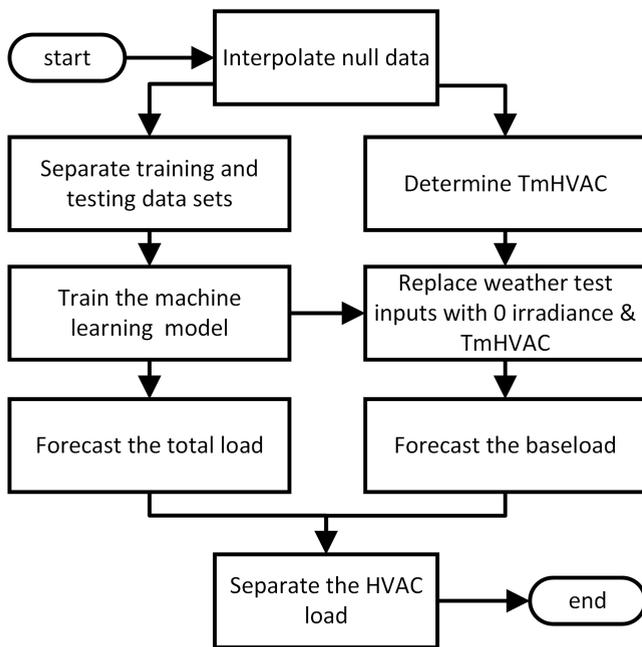


FIGURE 8. The step by step procedure of the novel HVAC separation for day ahead case (Sep FCST). Determining the temperature where the HVAC system first does not operate (TmHVAC) and replacing the weather inputs to be this TmHVAC and zero solar irradiance to forecast the baseload are very important steps to the method that distinguish it from other approaches.

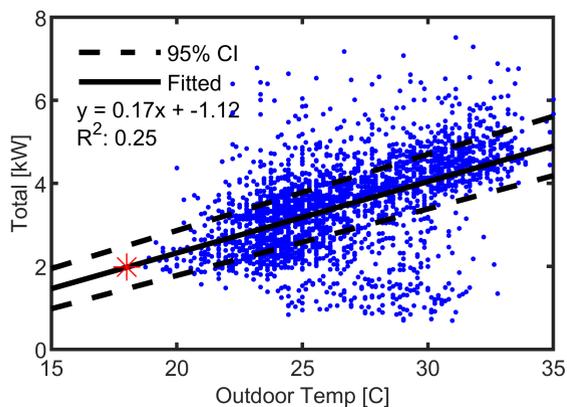


FIGURE 9. The V-type curve for the total residential energy usage of SHINES House #1 and the outdoor temperature corresponding to the 2020 hourly power data. A linear fit of the relationship is introduced and the TmHVAC reference variable of 18 C for the proposed method is marked by the red star.

previously described data driven LSTM model is not directly applicable to those homes.

To address this problem, this paper proposes a novel HVAC separation method from residential meter data. There are few other methods for HVAC load separation that are in the category of non-intrusive load monitoring approaches. Liang and Ma presented a non-homogeneous Factorial Hidden Markov Model (MN-FHMM) to dis-aggregate the HVAC load profile and to learn a residence HVAC patterns [23]. Song *et al.* proposed a time-frequency masking technique paired with a deep learning model to identify the HVAC load [26]. Both of these studies employ complex probability

and frequency domain signal analysis based calculations, which suffer from longer computational runtimes when implemented as computer programs; in addition, they were presented only conceptually, without emphasis on reproducibility.

Ming Liang *et al.* presented a daily average separation method based on the linear relationship between the outdoor temperature and HVAC operation as well as day type grouping [27]. They used a base power level without HVAC which is identified from mild day usage. It is subtracted from the average of hot and cold days to isolate the HVAC power, but only at the daily level without presenting a method for a higher resolution time-series, day-ahead forecasting. The novel method presented in this paper fills a gap in the literature by expanding the previous approaches to include time-series, day-ahead forecasting at the hourly resolution by utilizing deep learning models and weather-based relationships to separate the HVAC energy usage.

This method is developed using an encoder-decoder LSTM model applied to forecasting the house total and baseload, i.e., load from human behavior unrelated to weather. From these predictions the HVAC is estimated via subtraction for further use in the BTM DR schemes. The authors note that this approach can be improved further as machine learning technologies advance as other models can be applied to the steps described.

Fig. 7 shows a simplified diagram that explains the mathematical steps of the proposed method and an example week of HVAC separation from July 16th to the 22nd of 2020. This example week is satisfactory but not as accurate as the HVAC forecasts shown in Section V from the dedicated circuit based data driven model. A step by step procedure to obtain the day-ahead HVAC separation (Sep. FCST) shows the key steps to predicting the baseload in detail (Fig. 8). The advantage with this HVAC separation approach is that only a smart meter for the home is required, resulting in reduced cost.

The HVAC separation method exploits the linear relationship between the total power of a home (in [kW]) and the outdoor ambient temperature the building experiences. This relationship may be referred to as a “V” curve with the home power increasing towards both extremes of temperature from a minimum temperature that is often found near the middle of the curve in the more mild temperatures. It may then be concluded that the portion of the home load that increases with more extreme temperatures most likely corresponds to the HVAC load component [28]. This implication provides the basis of the cooling load separation method presented in this section.

The SHINES House #1 hourly temperature to total load relationship, the “V-Curve”, over the entire year of 2020 is shown in Fig. 9, and the temperature at which the HVAC load is at its minimum (TmHVAC) is indicated by a red star. It is important to note that the home is located in Florida, which has a hot and humid climate year-round. Therefore, the lower temperature half of this V-Curve, which corresponds to heating, no longer needs to be considered as the HVAC does

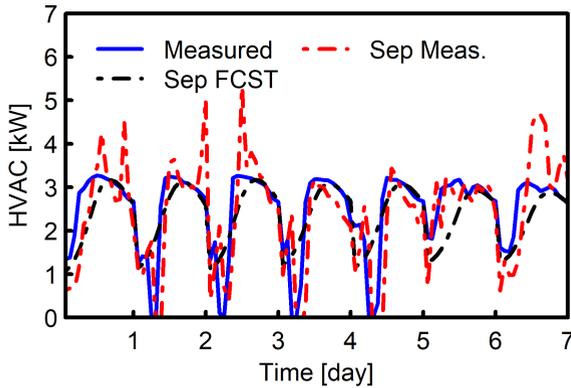


FIGURE 10. HVAC load separated from the total measured (Sep Meas.) and forecasted (Sep FCST) for the week of August 2, 2020 during the daylight hours, 6am to 9pm of SHINES House #1. This week is also shown in Fig. 4 with forecasts from measured data that outperform the separated method.

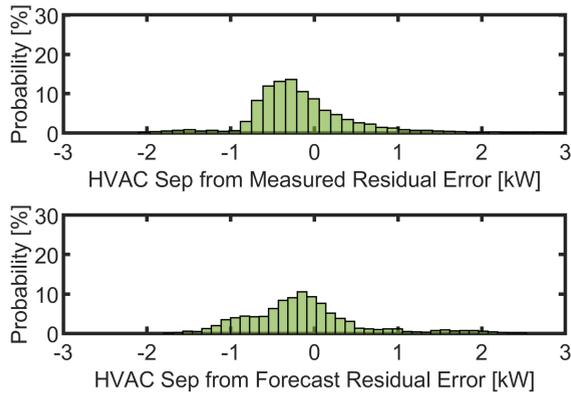


FIGURE 11. The approach to separate from the measured data (top) has approximately 40% probability of an error between 0.35 kW and 0.5 kW in the entire test window of over three months. The distribution shows that the results could be further improved by adjusting the selected TmHVAC reference variable for the baseload forecasts such that the errors are centered around 0 kW.

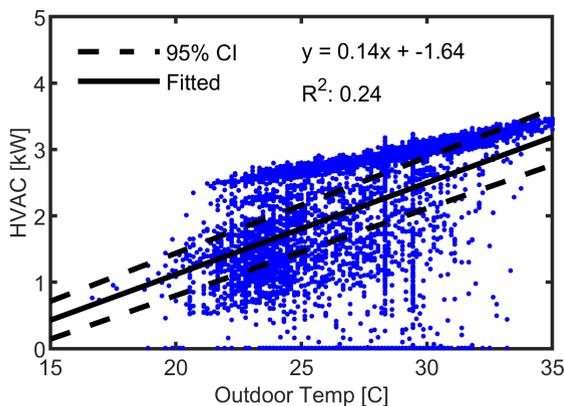


FIGURE 12. HVAC experimental data from the SHINES Home #1 for the years 2018 and 2019. A linear regression was completed to model the HVAC following traditional methods. This line is used in Table 2 to compare with the novel separation method.

TABLE 2. Summary of residential HVAC separation errors and comparison of linear model from data obtained by dedicated circuit separation traditional approach.

Hourly Metric	Sep Meas	Sep FCST	LR
Average Load [kW]	2.174	2.174	2.174
RMSE [kW]	0.639	0.699	0.772
CV(RMSE) [%]	29.4	32.2	35.5
R^2	0.445	0.335	0.189
Daily Metric	Sep Meas	Sep FCST	LR
Average Load [kWh/day]	51.670	51.670	51.670
RMSE [kWh/day]	5.728	6.957	8.481
CV(RMSE) [%]	11.1	13.5	16.25
R^2	0.665	0.506	0.1548

not operate in this range.

A linear fit of the total load against the temperature is used to model the HVAC system’s general performance as it is the largest weather dependent load in a home. The large variation outside the 95% confidence interval shown represents the other loads and human behavior shaped influences. Therefore, a comparison of different TmHVAC selections along the HVAC linear model can be made and compared depending on the coefficient of determination, R^2 . The conditions of the minimum HVAC load for the SHINES House #1 were determined to be at a temperature of 18 °C and a solar irradiance of 0 W/m².

For the data set from FL, this condition occurred naturally only at night. This observation is important as it provides insight on the level of influence for solar irradiance on the HVAC load, which may or may not contribute to the total house load. It is clear that, for times when the HVAC system does not operate, the total load of the home is comprised of other typical residential appliances and equipment, excluding the HVAC. The total load in this scenario is considered to be the “baseload” of the house for the proposed HVAC separation method.

Following the diagram in Fig. 7, after the value of TmHVAC is determined, the LSTM encoder-decoder model is employed to predict the baseload for the house. First, the LSTM model is trained normally on the total power for the home. Then, the temperature and irradiance inputs are replaced in the input file to the model during forecasting with the TmHVAC and 0 W/m² irradiance as if it were night and the HVAC system did not operate. With these modified inputs, the LSTM produces the “baseload”. Finally, to separate the HVAC load, the baseload is subtracted from the known total power of the home. For HVAC load predictions, the forecasted baseload can be subtracted from a day-ahead forecast for the entire total load, which may also be obtained from the LSTM model. This can serve in place of a forecast from an HVAC data driven model as it can be obtained 24 hours before the time of interest from smart meter data alone.

Both exercises were completed for the SHINES home (Fig. 10) and provided satisfactory results according to the ASHRAE building modeling standards of less than 35% CV (RMSE) at the hourly level and less than 10% at the daily resolution as described in the previous section (Table 2.)

In Table 2, a comparison is also provided against a linear regression model based on the increase of the HVAC load per degree of outdoor temperature in the summer, for which the data is shown in Fig. 12. This method is the traditional approach as a machine learning benchmark model for HVAC demand forecasting based on the use of dedicated equipment to separate the load [16].

Further analysis of the results show that over the summer, the probability distribution of the HVAC separation error is shifted to the left by a factor of 0.5 kW (Fig. 11). This indicates that, while the temperature selected is suitable, a more detailed investigation into the TmHVAC selection from the V-curve using a regression model may yield improved results. The HVAC separated load resulting from the forecasted total load (Sep FCST) does not experience any shifting from the center, which is due to the total load forecasts having an error distribution that varies between being both too high and too low.

Although the forecasts are satisfactory with an hourly and daily CV (RMSE) of 32.2% and 13.5%, respectively, they do not capture more irregular spikes caused by human behavior as well as measured data (Sep Meas.) which yields better results of an hourly CV(RMSE) of 29.4% and daily of 11.1%. The Sep FCST results yield great value however as they actually have slightly better accuracy than predicting the HVAC from the data obtained through dedicated measurement as presented in Section V, with an RMSE of .639 as opposed to 0.701 kW. This indicates that the added cost of dedicated circuit equipment is not supported by this study as opposed to the novel separation method presented which is sufficient from smart meter data alone.

VII. BEHIND-THE-METER CONTROLS AND CASE STUDIES FOR A NZE HOUSE

The presented BTM demand response controls, which may be deployed through HEMS, are enabled by the HVAC and PV energy forecasting methods proposed previously. The BTM control schemes target modern net zero energy (NZE) smart homes due to their increasing popularity, and it is assumed that such smart homes have a combined BES and PV system that is capable of storing and generating enough energy to maintain NZE status.

Because the SHINES experimental smart House #1 was not designed as a NZE home, both the measured data and the forecasts for the PV generation are scaled by a factor of 2.353 to make the actual, annual energy generated by the PV installation, 5705.961 kWh, equal to the total annual energy used by the home, 13426.607 kWh.

A general block diagram of the proposed BTM control approach is shown in Fig. 13. It includes five main components: processing of the experimental data, forecasting,

HVAC control, real time HEM, and monitoring. The raw data is processed before it can be used as the input to the LSTM model. This pre-processing is responsible for cleaning and normalization. The forecasting component from Fig. 13 uses LSTM models to forecast the residential load and PV generation for the next day. Its output is employed by the HVAC control component to schedule the HVAC operation via adjustments of the thermostat setpoint. For example, on a typical summer day, when there is surplus PV generation, the setpoint is set lower such that the HVAC system operates as energy storage. When the house power demand is high, the setpoint is increased in order to reduce the residential power. The HEM target is set when scheduling the HVAC operation based on the forecasting. A Battery Energy Storage (BES) system is incorporated in the real time HEM to eliminate variations caused by the difference between forecasting and real time data.

The proposed control approach uses a thermodynamic model. The HVAC thermal model parameters and the equivalent thermal resistance for the house of 4°C/kW (see Fig. 14) were calculated using the experimental data from June to September of 2018, 2019, and 2020 of the SHINES House #1. Moreover, only the data corresponding to outdoor temperatures higher than 33°C were considered since DR control is mostly likely to be employed only in such cases.

The HEMS operates under the assumption that homeowners may opt into HVAC load dispatch controls on any given day. Under this assumption, the homeowner is presented with the optimized schedule for day-ahead HVAC load that minimizes the electricity cost. The impact of the HVAC setpoint control and battery usage on the total energy purchased from the grid is analyzed using a cost-calculation based on time-of-use pricing (TOU) from [29]. In this scheme, the most expensive time of day to use electricity is 4pm-9pm at \$0.43 instead of \$0.27 per kWh outside this time-frame.

In the remainder of this section, two computational case studies of HVAC controls implemented as described in Fig. 13 are presented in order to assess the impact of HVAC control on the net power flow of a modern NZE smart home equipped with PV solar generation and batteries with different sizes. The LSTM based models presented in previous sections are used to forecast the HVAC load and PV generation. These predictions, together with the measured data (used for comparison purposes), are used as an input into the P_{Net} equation for HVAC demand response under two scenarios: Scenario 1, where one battery is used and Scenario 2, where two batteries are used. Battery sizes and power ratings are assumed to be similar to common industrial batteries such as the Tesla Powerwall with 13.5kWh of usable capacity and 5kW power rating [30].

Furthermore, to reduce the cost to the user in both scenarios, the batteries are assumed to be charged at the start of the day and began its operation at 6am to hold the house's net power flow at 0kW. In this approach, the battery discharges and recharges by the excess generation in the middle of the day as shown in Fig. 16. The setpoint is also selected as lower

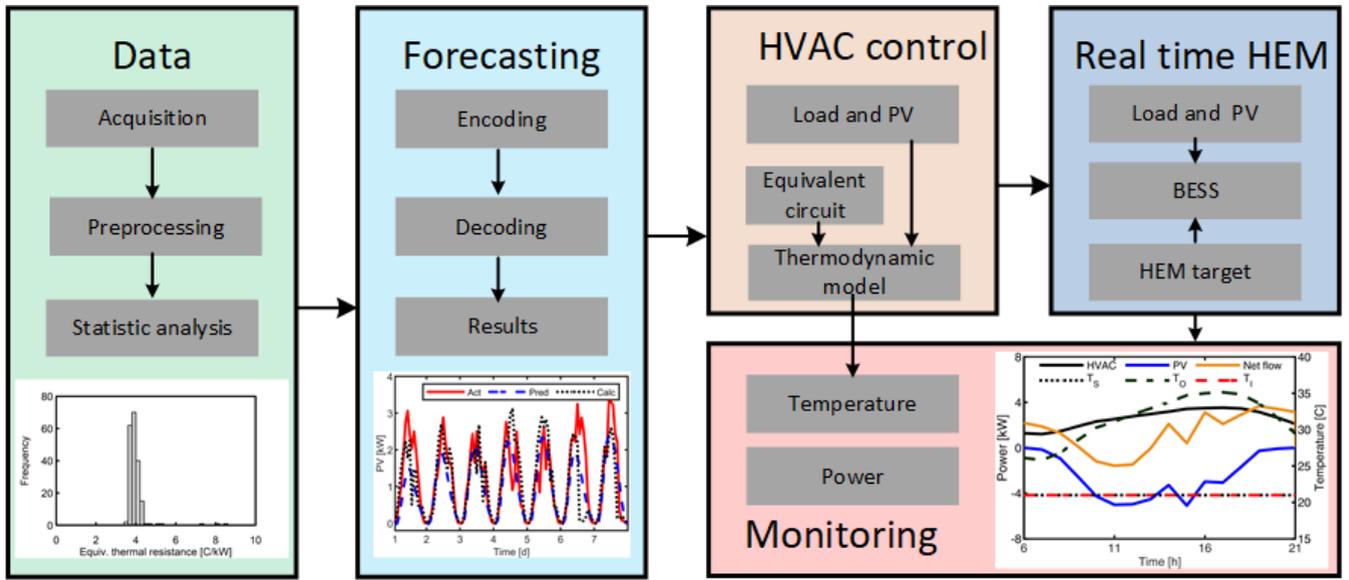


FIGURE 13. Flow diagram for the proposed HEM control. The HVAC control is planned one day ahead based on the forecasting results. Additionally, the BESS is controlled real time via the HEMS to meet the control target and compensate for total load, based on measured and forecasted data, as needed.

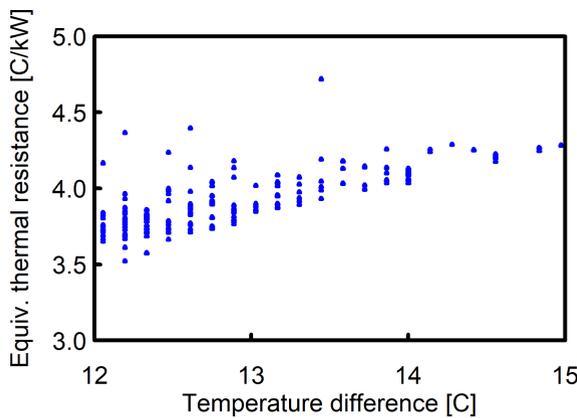


FIGURE 14. Calculated equivalent thermal resistance based on daytime data from June to September 2018–2020. Only data points with a difference greater than or equal to 12 °C between outside and inside ambient temperature were considered to correspond to a typical mid-day circumstance in which controls should be applied.

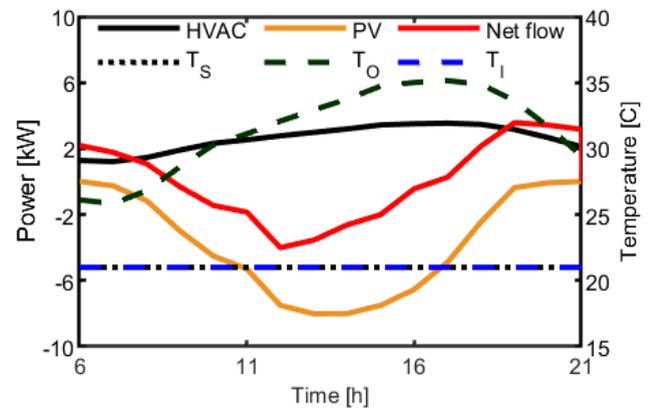


FIGURE 15. Representative plots for SHINES House #1 on July 12, 2020. The house is transformed into a NZE house via scaling with a factor of 2.353 for the annual PV generation and total energy use.

than the 21°C from the reference case to use the available PV generation and pre-cool the home, which relies on the home’s thermal inertia to reduce necessary cooling later in the day. For example, the plots in Fig. 15 show the measured HVAC power, PV generation, and NZE net power flow along with the outdoor and indoor temperatures on July 11, 2020, as a typical summer day in Pensacola, FL (location of SHINES House #1). In this reference case, the indoor temperature was considered to be constant at 21°C. The HVAC power is consistently high during the afternoon and peaks around 5pm with the outdoor temperature. In this uncontrolled state, the user needed to pay higher prices during the evening to cool their house, and their total cost of electricity from TOU prices was \$5.73.

The proposed control approach was applied in Scenario 1 to House #1 equipped with one battery of 13.75 kWh capacity. The plots from Fig.17 show the power flows for a representative summer day, where the measured data resulted in a net zero power flow for the day. The forecasted variables indicated that the battery would start discharging in the evening slightly earlier and would, thus, deplete its charge before the entire evening load was negated. This load scheduling comparison shows that even very good forecasts can cause small discrepancies in day-ahead load scheduling and cost predictions.

In Scenario 2, where the house is equipped with 27 kWh battery capacity (achieved through two separate batteries of capacity as in Scenario 1), the HVAC dispatch control based HEMS can maintain a net zero power flow through the entire high cost period of the day from 4-9pm in both

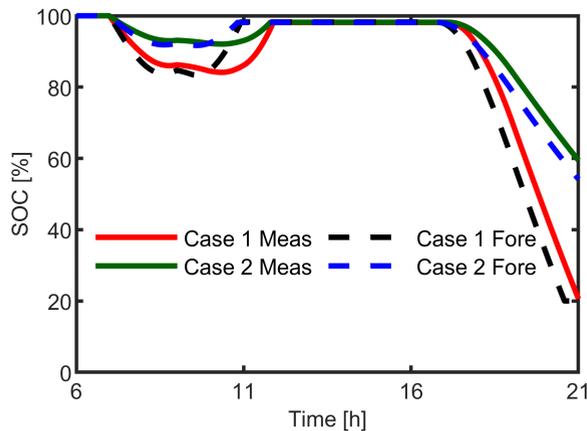


FIGURE 16. Battery SOC over both Scenario 1 and Scenario 2. The batteries discharge in the early morning and recharge during times of excess PV generation.

TABLE 3. Consumer Benefit Comparison for Scenario 1 and 2.

Scenario 1 Daily Changes	Measured	Forecasted
Cost Savings [\$]	2.85	2.22
HVAC Energy Increase [kW]	4.968	4.968
Total Energy Purchased (4-9pm) [kW]	0.000	1.401
Scenario 2 Daily Changes	Measured	Forecasted
Cost Savings [\$]	2.77	2.40
HVAC Energy Increase [kW]	4.744	4.744
Total Energy Purchased (4-9pm) [kW]	0.000	0.000

the forecasted and measured cases (Fig. 18). However, the additional battery capacity is not fully utilized, as the battery has more than 60% capacity remaining. Also, because of its higher capacity, the battery incurs a higher cost to charge to full capacity (during night time), which, in turn, results in fewer savings. Therefore, there is less incentive for the homeowner to purchase an additional battery. However, a larger battery capacity could be utilized as additional backup supply for days with poor PV generation, e.g., cloudy days.

Both scenarios indicate that shifting the HVAC setpoint to a value such that the home pre-cools when the user is assumed to be absent from the home can save the homeowner more than \$2 per day as reported in Table 3. The proposed HVAC control scheme - which operates the system to act as a load bank for otherwise unused PV generation - increases the total amount of energy used for air conditioning, but, the environmental impact is lessened by the reduced need for electricity from the grid, which often heavily relies on fossil fuels.

VIII. CONCLUSION

An encoder-decoder LSTM ML model was developed for hourly, day-ahead forecasts of residential HVAC usage and PV generation based on previous load usage and future weather inputs in place of a weather forecasts. These predictions utilize experimental 15-minute resolution data aggregated to the 1 hr timestep. The respective goodness of fit for

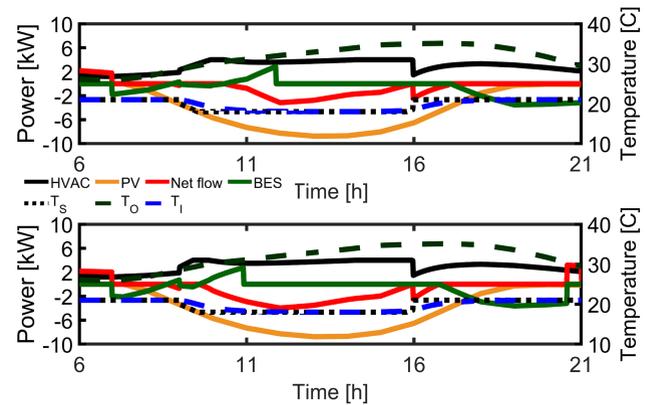


FIGURE 17. Scenario 1: day-ahead load shifting schedule for HVAC based on PV generation (measured top, forecasted bottom) with 13.5 kWh of battery capacity. The forecast of PV ends slightly earlier than the measured data thus causing the positive net power flow at night.

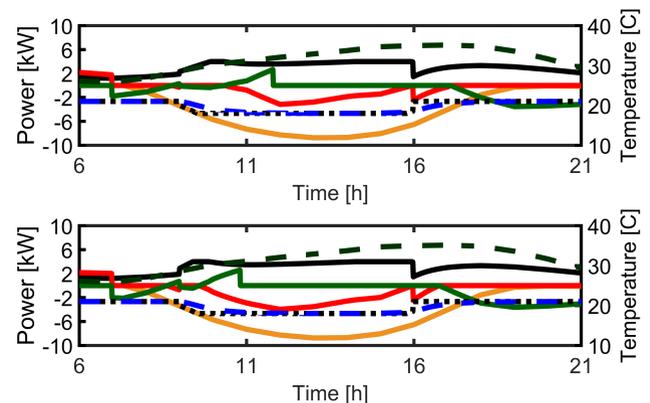


FIGURE 18. Scenario 2: Both the measured PV (top) and forecasted PV (bottom) based HVAC controls result in 0 kWh purchased from the grid during the day with a 27 kWh battery. The only demand-related cost to the user in a 24 hour period is for recharging the battery at night when ToU rates are cheaper.

the PV and HVAC models were 0.975 and 0.486 at the hourly level.

The presented method for the forecast of the residential building total load and separation of the HVAC load, which is typically a residence's largest load component, offers a significant advantage in that it does not require a dedicated circuit for HVAC monitoring. The technique only requires data that is made available from smart meters and a weather forecast, which makes it suitable for large scale field deployment.

Furthermore, the method utilizes non-intrusive load monitoring with smart meters at the house level that can enable advanced BTM controls for HVAC systems, which is an especially notable achievement. It was shown that the encoder-decoder LSTM-based HVAC separation method produced excellent results when compared to the ASHRAE building model standards as applied to HVAC systems. A daily CV(RMSE) of 11.1% and 13.5% were found by employing measured and forecasted total load data, respectively.

The technique for separation is also versatile as other ma-

chine learning models that have been optimized, in addition to the LSTM model encoder-decoder model, may be applied to improve the results of separation even further. Future research will include further validation of the proposed techniques by considering a wide variety of house types, climate zones, occupancy, and load patterns, i.e., occupant behavior.

Case studies were also performed in this paper to illustrate the application of the day-ahead forecasts in a BTM control approach with the objective of operating a smart house such that it is considered NZE. The proposed control approach was also utilized to conduct a BTM user-centric cost analysis that indicated the possibility of cost savings under realistic TOU pricing.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (NSF) grant ECCF 1936494 and the NSF Graduate Research Fellowship under Grant No. 1839289. The support of the a Department of Education (DoEd) GAANN Fellowship is also gratefully acknowledged. Any findings and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the NSF and DoEd.

REFERENCES

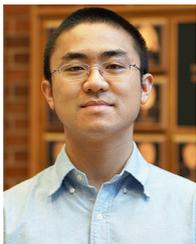
- [1] H. Gong, E. S. Jones, R. E. Alden, A. G. Frye, D. Colliver, and D. M. Ionel, "Demand Response of HVACs in Large Residential Communities Based on Experimental Developments," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2020, pp. 4545–4548.
- [2] A. Bandyopadhyay, J. P. Conger, and M. E. Webber, "Energetic potential for demand response in detached single family homes in austin, tx," in *2019 IEEE Texas Power and Energy Conference (TPEC)*, 2019, pp. 1–6.
- [3] W. Tumin, M. M. Olama, and S. M. Djouadi, "Adaptive control for residential hvac systems to support grid services," in *2021 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2021, pp. 01–05.
- [4] K. Kurte, J. Munk, O. Kotevska, K. Amasyali, R. Smith, E. McKee, Y. Du, B. Cui, T. Kuruganti, and H. Zandi, "Evaluating the adaptability of reinforcement learning based hvac control for residential houses," *Sustainability*, vol. 12, no. 18, 2020. [Online]. Available: <https://www.mdpi.com/2071-1050/12/18/7727>
- [5] F. Hafiz, M. A. Awal, A. R. d. Queiroz, and I. Husain, "Real-Time Stochastic Optimization of Energy Storage Management Using Deep Learning-Based Forecasts for Residential PV Applications," *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 2216–2226, 2020.
- [6] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an lstm neural network and synthetic weather forecast," *IEEE Access*, vol. 8, pp. 172 524–172 533, 2020.
- [7] M. Borowski and K. Zwolińska, "Prediction of cooling energy consumption in hotel building using machine learning techniques," *Energies*, vol. 13, no. 23, 2020. [Online]. Available: <https://www.mdpi.com/1996-1073/13/23/6226>
- [8] Y. Lu, Z. Tian, R. Zhou, and W. Liu, "Multi-step-ahead prediction of thermal load in regional energy system using deep learning method," *Energy and Buildings*, vol. 233, p. 110658, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378778820334447>
- [9] R. Sethi and J. Kleissl, "Comparison of short-term load forecasting techniques," in *2020 IEEE Conference on Technologies for Sustainability (SusTech)*, 2020, pp. 1–6.
- [10] C. H. Kim, M. Kim, and Y. J. Song, "Sequence-to-sequence deep learning model for building energy consumption prediction with dynamic simulation modeling," *Journal of Building Engineering*, vol. 43, p. 102577, 2021.
- [11] M. Razghandi and D. Turgut, "Residential appliance-level load forecasting with deep learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.
- [12] R. E. Alden, H. Gong, C. Ababei, and D. M. Ionel, "Lstm forecasts for smart home electricity usage," in *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*, 2020, pp. 434–438.
- [13] D. Syed, H. Abu-Rub, A. Ghayeb, and S. S. Refaat, "Household-level energy forecasting in smart buildings using a novel hybrid deep learning model," *IEEE Access*, vol. 9, pp. 33 498–33 511, 2021.
- [14] R. Sendra-Arraz and A. Gutiérrez, "A long short-term memory artificial neural network to predict daily hvac consumption in buildings," *Energy and Buildings*, vol. 216, p. 109952, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378778819312861>
- [15] G.-F. Xia, Z.-Q. Tian, and Z.-K. OuYang, "Research on forecast of daily electricity consumption of household air conditioning based on improved long-short memory network," in *Journal of Physics: Conference Series*, vol. 1948, no. 1. IOP Publishing, 2021, p. 012018.
- [16] K. Zingre, S. Srinivasan, and M. Marzband, "Cooling load estimation using machine learning techniques," 2019.
- [17] Y. Ma, Y. Tang, B. Li, and B. Qi, "Residential high-power load prediction based on optimized lstm network," in *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, 2020, pp. 538–541.
- [18] L. Xudong, L. Shuo, and F. Qingwu, "Prediction of building heating and cooling load based on ipso-lstm neural network," in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 1085–1090.
- [19] Electric Power Research Institute. (2020) DOE SHINES Residential Demonstration. [Online]. Available: <https://dashboards.epri.com/shines-residential/dashboard>
- [20] Jason Brownlee, "Long Short-Term Memory Networks With Python - Develop Sequence Prediction Models With Deep Learning," *Machine Learning Mastery Ebook*, Ed. v1.6, 2019.
- [21] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," [Online]. Available: <https://arxiv.org/abs/1406.1078>, 2014.
- [22] M. Manivannan, B. Najafi, and F. Rinaldi, "Machine Learning-Based Short-Term Prediction of Air-Conditioning Load through Smart Meter Analytics," *Energies*, vol. 10, p. 1905, 11 2017.
- [23] H. Liang and J. Ma, "Separation of Residential Space Cooling Usage From Smart Meter Data," *IEEE Transactions on Smart Grid*, vol. PP, pp. 1–1, 01 2020.
- [24] G. R. Ruiz and C. F. Bandera, "Validation of calibrated energy models: Common errors," *Energies*, vol. 10, no. 10, p. 1587, 2017.
- [25] B. Bezyan and R. Zmeureanu, "Machine learning for benchmarking models of heating energy demand of houses in northern canada," *Energies*, vol. 13, no. 5, p. 1158, 2020.
- [26] J. Song, Y. Lee, and E. Hwang, "Time-frequency mask estimation based on deep neural network for flexible load disaggregation in buildings," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3242–3251, 2021.
- [27] M. Liang, Y. Meng, N. Lu, D. Lubkeman, and A. Kling, "Hvac load disaggregation using low-resolution smart meter data," in *2019 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2019, pp. 1–5.
- [28] Q. Meng, M. Mourshed, and S. Wei, "Going beyond the mean: Distributional degree-day base temperatures for building energy analytics using change point quantile regression," *IEEE Access*, vol. 6, pp. 39 532–39 540, 2018.
- [29] S. C. E. E. for What's Ahead, "Time-of-use (tou) rate plans," <https://www.sce.com/residential/rates/Time-Of-Use-Residential-Rate-Plans>.
- [30] Tesla, "Powerwall battery datasheet," https://www.tesla.com/sites/default/files/pdfs/powerwall/Powerwall_2_AC_Datasheet_EN_NA.pdf.



ROSEMARY E. ALDEN (Graduate Student Member, IEEE) received in 2021 her B.S. in electrical engineering from the University of Kentucky, Lexington, Kentucky, USA, where she is currently a PhD student in the SPARK Lab. During her undergraduate studies, she was an L. Stanley Pigman Scholar and an IEEE PES Plus Scholarship recipient. She is now a National Science Foundation (NSF) Graduate Research Fellow. She was an intern with the Electric Power Research Institute (EPRI) and with the National Renewable Energy Laboratory (NREL). She has been serving as the inaugural chair for the joint student chapter of IEEE PES and IAS at UK. Her research interests include machine learning, electric load forecasting, smart buildings and grid, and power system grid analysis for renewable energy integration.



CRISTINEL ABABEI (M'00-SM'14) received the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2004 and the M.Sc. in signal processing and B.S. in microelectronics degrees from the Technical University "Gh. Asachi" of Iasi, Romania. He is an associate professor in the Dept. of Electrical and Computer Engineering (ECE), Marquette Univ. Prior to that, from 2012 to 2013, he was an assistant professor in the Dept. of Electrical Engineering, State University of New York (SUNY) at Buffalo. Between 2008 to 2012, he was an assistant professor in the Dept. of ECE, North Dakota State University (NDSU). From 2004 to 2008, he worked as an R&D engineer for Magma Design Automation, San Jose, CA. His current research interests include machine learning techniques applied to energy optimization in multi-core processors and datacenters, co-simulation of buildings and distribution networks, embedded systems design for environment monitoring, and IoT security.



HUANGJIE GONG (S'18) is a Ph.D. candidate in the SPARK Laboratory, Electrical and Computer Department at University of Kentucky (UK). He started his studies at UK in 2017 after receiving his M.S. degree in Control Theory and Engineering from Southwest Jiaotong University, Chengdu, China. At UK, he has been working as a Teaching Assistant and as a Research Assistant on projects sponsored by DOE, NSF, companies, utilities, including TVA and LG&E-KU, and EPRI. He is the main developer of an HPC-based large-scale co-simulation software framework for energy use in buildings and power flow in electric power distribution systems incorporating EnergyPlus and OpenDSS as computational engines. Huangjie also conducts research on machine and deep learning for human behavior and energy management in smart buildings. Additional research topics include HVAC, water heater, EV, PV, and battery modelling and control, power system controls and protection, net-zero-energy (NZE) buildings, aggregate models for loads and DER in electric power distribution systems, and smart grids.



DAN M. IONEL (M'91-SM'01-F'13) received the M.Eng. and Ph.D. degrees in electrical engineering from the Polytechnic University of Bucharest, Bucharest, Romania. His doctoral program included a Leverhulme Visiting Fellowship with the University of Bath, Bath, U.K and he later was a Postdoctoral Researcher with the SPEED Laboratory, University of Glasgow, Glasgow, U.K.

Dr. Ionel is a professor of electrical engineering and the L. Stanley Pigman Chair in Power with the University of Kentucky, Lexington, KY, USA, where he is also the Director of the Power and Energy Institute of Kentucky and of the SPARK Laboratory. He previously worked in industry for more than 20 years. Dr. Ionel's current research group projects on smart grid and buildings, and integration of distributed renewable energy resources and energy storage in the electric power systems are sponsored by NSF, DOE, industry and utilities. He published more than 200 technical papers, including some that received IEEE awards, was granted more than 30 patents, and is co-author and co-editor of the book "Renewable Energy Devices and Systems – Simulations with MATLAB and ANSYS", CRC Press.

Dr. Ionel received the IEEE PES Veinott Award, was the Inaugural Chair of the IEEE IAS Renewable and Sustainable Energy Conversion Systems Committee, an Editor for the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, and the Technical Program Chair for IEEE ECCE 2016. He is the Editor in-Chief for the Electric Power Components and Systems Journal, and the Chair of the Steering Committee for IEEE IEMDC.



EVAN S. JONES (Graduate Student Member, IEEE) is a Ph.D. student and a GAANN Graduate Fellow in the SPARK Laboratory at the University of Kentucky (UK), Lexington, KY, USA. He received two B.S. degrees in electrical engineering and computer engineering with a minor in computer science, and an undergraduate certificate for studies in power and energy from UK in 2019. He has also served as the vice president for the joint student chapter of IEEE PES and IAS at UK. In 2020, he was awarded a multi-year GAANN Ph.D. Fellowship administered by the U.S. Department of Education. His research focuses on building energy models (BEMs), virtual power plants (VPPs), electric power systems, and renewable energy generation and integration.

...