Digital Twin for HVAC Load and Energy Storage based on a Hybrid ML Model with CTA-2045 Controls Capability

Rosemary E. Alden
SPARK Lab, ECE Department
University of Kentucky
Lexington, KY USA
rosemary.alden@uky.edu

Evan S. Jones
SPARK Lab, ECE Department
University of Kentucky
Lexington, KY USA
sevanjones@uky.edu

Steven B. Poore
SPARK Lab, ECE Department
University of Kentucky
Lexington, KY USA
stevenpoore@uky.edu

Huangjie Gong
SPARK Lab, ECE Department
University of Kentucky
Lexington, KY USA
huangjie.gong@uky.edu

Abdullah Al Hadi
SPARK Lab, ECE Department
University of Kentucky
Lexington, KY USA
abdullah.hadi@uky.edu

Dan M. Ionel
SPARK Lab, ECE Department
University of Kentucky
Lexington, KY USA
dan.ionel@ieee.org

***Abstract*—Building modeling, specifically heating, ventilation, and air conditioning (HVAC) load and equivalent energy storage calculations, represent a key focus for decarbonization of buildings and smart grid controls. Widely used white box models, due to their complexity, are too computationally intensive to be employed in high resolution distributed energy resources (DER) platforms without simulation time delays. In this paper, an ultra-fast one-minute resolution Hybrid Machine Learning Model (HMLM) is proposed as part of a novel procedure to replicate white box models as an alternative to wide spread experimental big data collection. Synthetic output data from experimentally calibrated EnergyPlus models for three existing smart homes managed by the Tennessee Valley Authority is used. The HMLM employs combined k-means clustering and multiple linear regression (MLR) models to predict minutely HVAC power with satisfactory nRMSE error of less than 10% across an entire year test set. An approach is provided to characterize HVAC systems through the newly proposed hybrid model as a generalized storage (GES) device suitable for DER control and event types in accordance with the Communication Technology Association (CTA) 2045 standard and Energy Star metrics such as "energy take", currently developed by industry, to unify household appliance controls.**

***Index Terms*—Battery Energy Storage System (BESS), Heating Ventilation and Air Conditioning (HVAC), Energy Storage, ANSI/CTA-2045-B, Energy Star, Energy Take, Home Energy Management (HEM), Demand Response (DR), machine learning, smart homes, smart grid**

I. Introduction

The heating, ventilation, and air conditioning (HVAC) system is an important component for building decarbonization and modeling as it is the largest residential building load component according to the US Energy Information Association (EIA). To evaluate the HVAC energy performance, there are three different methods: a physics-based model known as a "white box", a statistical or data driven model called a "black box", and a hybrid model that combines both white and black box known as a "gray box" model.

For white box modeling, there are many detailed characteristics required, such as the type of HVAC system, SEER rating, and building characteristics e.g., insulation, air flow rate, ventilation, indoor/outdoor climatic conditions, door and window types, size/area, etc. Due to the number of parameters and multi-physics equations involved, these models are time consuming to develop and simulate [1]. Black box models may overcome many drawbacks of other models as they can be derived from big data as statistical and fast data driven models. Their integration into co-simulation platforms and model-in-the-loop calculations enables building model analysis at faster speeds such as the real-time optimization of energy storage (ES) in [2]. Some researchers have also recently proposed white-box models to create synthetic data sets to train black models [3], [4].

This paper brings further contributions through the deployment of a new Hybrid ML Model (HMLM) to replicate building HVAC power usage from synthetic output data of a calibrated EnergyPlus model, a widely used white-box model [5]. High resolution demand response (DR) and HVAC case studies are also included using the HMLMs. The HVAC status OFF, indoor temperature (T) is modeled and paired with a physics based equation.

*Huangjie Gong and Abdullah Al Hadi have been with the SPARK Lab at the University of Kentucky and are now with ABB Corporate Research Center, Raleigh, NC, USA and Enginuity Power Systems, Clinton Twp, MI, USA, respectively.

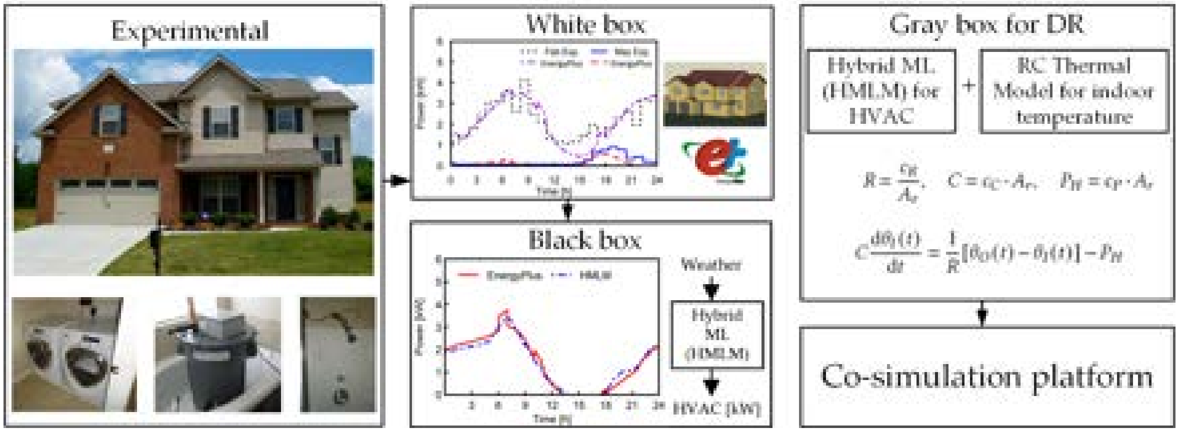

Fig. 1. Experimental data and building characteristics were used to train a white box model, EnergyPlus-based digital twin. The synthetic output data was used to train a black box ML model which can be used in a gray box model with co-simulation platforms to greatly reduce simulation time compared with the original white box model.

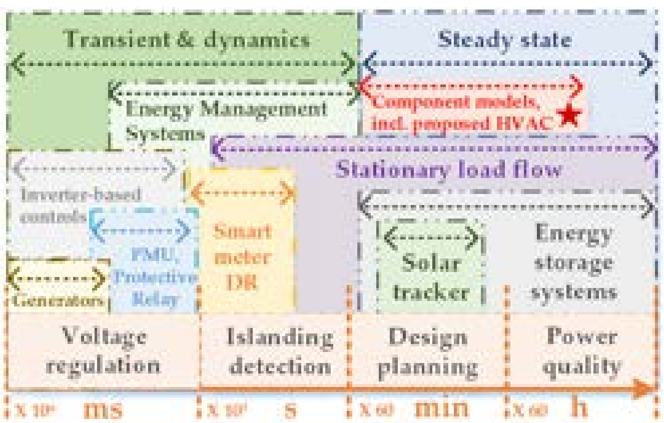

Fig. 2. Time scale comparison for electric power systems operation and control. The proposed HVAC model fills in the gap of ultra-fast multi-physics simulations with a one minute time resolution, as marked by the red star.

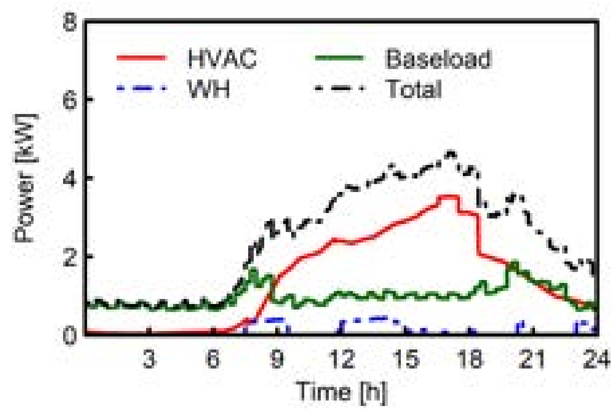

Fig. 4. Total synthetic load for a home based on the proposed HMLM model of a conventional HVAC system with high load, typical experimental baseload, and an high efficiency Heat Pump Electric Water Heater (HPEWH).

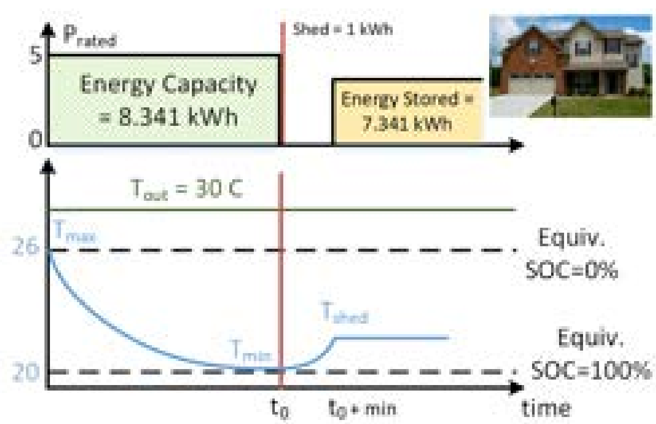

Fig. 3. Proposed application of the CTA-2045 for the thermal energy shed command in HVAC systems. Energy capacity for an example experimental conventional home is shown for a variable speed, 13 SEER system.

II. Smart Home Simulation with the Newly Proposed Model and CTA-2045

In electric power system modeling, data of varying time resolution is needed depending on the equipment used for controls and simulations (Fig. 2). Machine learning (ML) models are applicable in many power system simulation scenarios, as they can be trained at various resolutions. They are typically employed for stationary load flow forecasts and complement the range of tools used from the micro-second to minute resolution with home energy management (HEM) system operations.

To implement HVAC controls in HEM systems, unification of GES modeling with industry standard communication protocols, such as those from the Consumer Technology Association (CTA) and Energy Star, is necessary so that batteries, water heaters, appliances, and now HVAC systems may receive the same signals [6]. CTA standard 2045 communication protocol indicates “Load-Up” and “Shed” commands [7] that

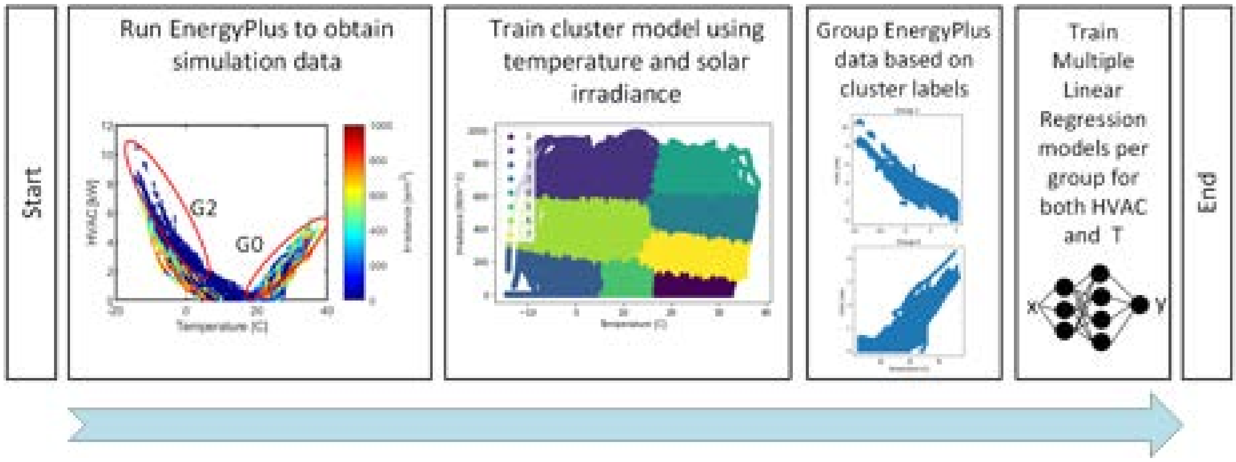

Fig. 5. Architecture of the Hybrid ML Model (HMLM) with k-means clustering and MLR using EnergyPlus simulation output data. The approximately linear trends of the HVAC system and outdoor temperature over the course of the year are isolated to improve performance of HVAC models across seasons and different types of days.

can be paired with "energy take" and equivalent state-of-charge (SOC) calculations to control HVAC systems as GES devices [8]. This creates a need to estimate the thermal energy storage in kWh and energy take of a home at a given time based on the current indoor T such that a new temperature setpoint can be selected to respond to the DR event signal (Fig. 3).

An example load profiles for a smart home based on synthetic and experimental data that could be used to test the impact of CTA-2045 derived DR and HVAC controls are shown in Fig. 4. The HMLM model proposed in this paper is used for the HVAC load following a calibrated EnergyPlus model of Tennessee Valley Authority (TVA) robotic houses [9]. The baseload and water heater profile data is publicly available from the DOE SHINES smart home project in Florida [10].

III. Hybrid ML Model and Digital Twin Description

The Hybrid Machine Learning Model (HMLM) proposed in this paper combines (1) k-means clustering and (2) multiple linear regression (MLR) models to create fast black box models of a more complex white box model in EnergyPlus. The output HVAC system data is a synthetic data set for the experimental home treated as representative based on its building characteristics and experimentally based calibration process, as previously stated.

In Fig. 5, the V-curve of HVAC power by temperature visualizes example approximately linear HVAC groupings caused by different weather conditions for which separate MLR models are trained. First, the k-means clustering is performed on input weather parameters that have the most influence over the HVAC load linearity, i.e. outdoor temperature and solar irradiance over an entire year. Different subsets of the inputs are labelled to isolate circumstances under which the HVAC system of a home would operate similarly, such as hot bright days in the summer, mild days in the shoulder months, and colder dark days in the winter. A k-value of eight was determined through numerical experimentation as a sufficient group size for separating approximately linear HVAC patterns with outdoor temperature.

Utilizing the group labels from the clustering, MLR models are trained and saved for use with their respective clusters. An input parameter study was completed to determine the best input structure for use in the MLR portion of the model (Table I). The considered inputs include outdoor temperature [°C], setpoint temperature [°C], the difference between the outdoor temperature and the indoor setpoint, the relative humidity [%], and solar irradiance [W/m^2]. The setpoint was assumed equal to the indoor temperature from the calibrated EnergyPlus model. In this study, the EnergyPlus model was simulated twice to create the training and test sets across two years. A typical meteorological year (TMY) weather was used during training to capture trends of historical performance. The different combinations had comparable, satisfactory results with less than 10% of nRMSE. The thermal inertia had only slight influence at the minutely resolution in comparison to the expected large influence at lower resolutions.

For use in DR case studies into HVAC controls, a gray box model is proposed using the HMLM and equivalent thermal resistance of a home. This gray box model can serve as a model-in-the-loop inside co-simulation platforms (Fig. 6) such as proposed in [11]. Parameters for the equivalent thermal model were calculated using a thermal envelope area and capacitance of 354 $°C \cdot m^2/kW$ and 0.011 $kWh/(°C \cdot m^2)$, respectively. The heat transfer function used in the RC thermal model is described as follows:

$$R = \frac{c_R}{A_r}, \quad C = c_C \cdot A_r, \quad P_H = c_P \cdot A_r, \tag{1}$$

$$C\frac{\mathrm{d}\theta_I(t)}{\mathrm{d}t} = \frac{1}{R}[\theta_O(t) - \theta_I(t)] - P_H, \tag{2}$$

where R, is the thermal resistance; C, the thermal capacitance;

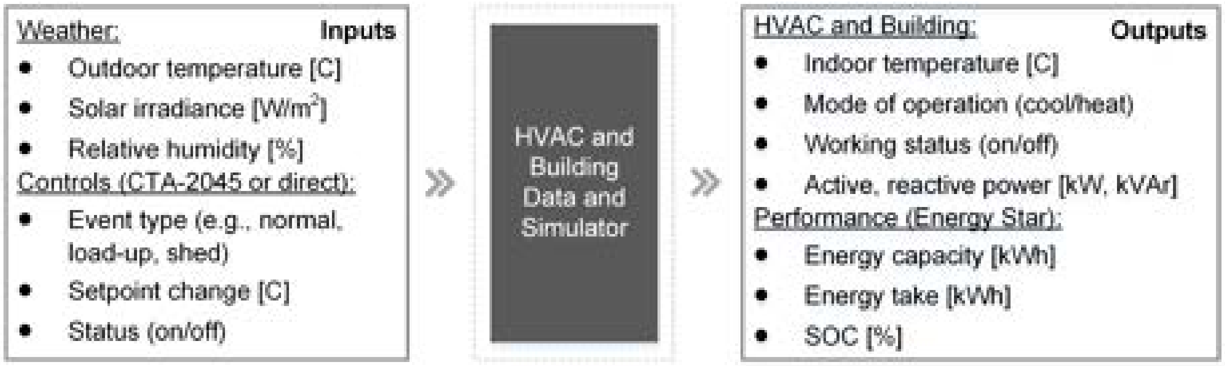

Fig. 6. Proposed co-simulation platform for smart home and component modeling following CTA-2045 and Energy Star industry standards. The platform is structured so that models are interchangeable for HVAC and other load components to increase simulation ease of adaptability.

Table I
CASE STUDY FOR A CONVENTIONAL HOME LOCATED IN KNOXVILLE, TN WITH A MINUTELY MAXIMUM HVAC POWER OF 4.7 AND 7.7 KW IN THE SUMMER AND WINTER.

Inputs	MAE [kW]	RMSE [kW]	nRMSE [%]	R^2 [-]
T_o, G, RH	0.348	0.486	6.3	0.820
T_o, T_d, G, RH	0.359	0.498	6.5	0.812
T_o, T_i, G, RH, T_p	0.321	0.452	5.9	0.844
T_d, G, RH, T_p	0.336	0.469	6.1	0.832

Table II
EXAMPLE HYBRID-MACHINE LEARNING MODELS OF ENERGYPLUS SYNTHETIC DATA FOR THREE BUILDING TYPES. THE INPUTS WERE T_o, T_d, G, AND RH FOR EACH MODEL.

Home Type	MAE [kW]	RMSE [kW]	nRMSE [%]	R^2 [-]
Conventional	0.359	0.498	6.5	0.81
Retrofit	0.125	0.173	3.5	0.88
NNZE	0.194	0.286	7.7	0.68

T_o = Outdoor T, T_i = Indoor T, G = Irradiance, T_d = T difference, T_p = T input at previous time, 15 to 60 min

P_H, heat transfer rate; θ_I is the indoor temperature; and θ_O, the outdoor temperature.

The RC model uses the thermal envelope and resistance of the home to calculate the change in indoor temperature during a DR event to turn the HVAC status off. After the DR event time window, a recovery period was assumed to return the temperature to the setpoint, during which the HVAC unit operates at rated power, until the same amount of energy is removed from the air as it took to raise the temperature. An example DR computational study of HVAC controls using a summer day illustrates how HVAC load is shed based on the combined gray box modeling (Fig. 7).

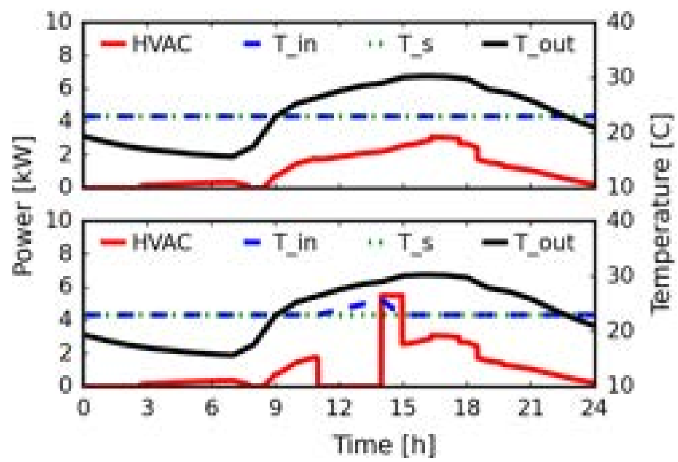

Fig. 7. Example use of the proposed machine learning HVAC and indoor temperature models during a DR event using experimental summer weather data from TN.

IV. CASE STUDY FOR CONVENTIONAL, RETROFIT, AND NNZE HOMES

The data used in this study is from experimentally validated EnergyPlus models of Tennessee Valley Authority (TVA) robotic field demonstration homes in Knoxville, [9]. Included in the study are three homes of conventional, retrofit, and Near Net Zero Energy (NNZE) type. The net annual energy use of the three homes is approximately 20, 12, and about 6 MWh respectively as there are significant differences in the construction and heat pump HVAC systems, i.e. in SEER rating, operational speed, and insulation type.

The HMLM model proposed in this paper was trained to provide minutely HVAC power for each home type from inputs of outdoor temperature (T_o), difference between outdoor and setpoint (T_d), relative humidity (RH), and irradiance (G) at the current minute only. The retrofit home with the most efficient HVAC system was modeled with the highest accuracy of the three homes with an R^2 of 0.88 and a nRMSE of 3.5% (Table. II). The residual error distributions for the homes in Fig. 9 are strongly cluster around zero, with up to 80% of all errors in the test year within ±0.25 kW.

Example days in the summer show the HMLM model captures the minutely trends of the HVAC power for each home (Fig. 8). Residential PV generation from the DOE SHINES field demonstration of two smart homes shows high

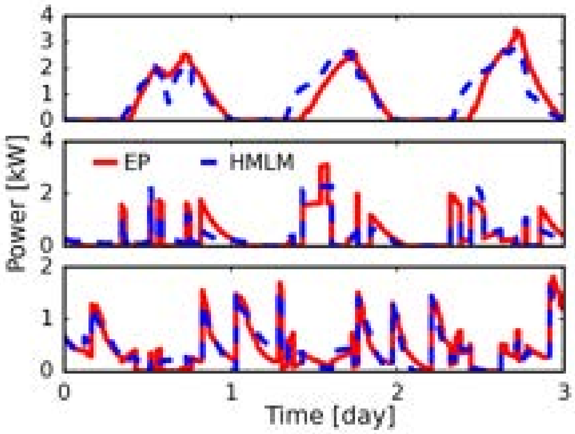

Fig. 8. Summer example HVAC calculations from the HMLM model of EP synthetic data for three days in June of the conventional (a), retrofit (b), and NNZE (c) homes.

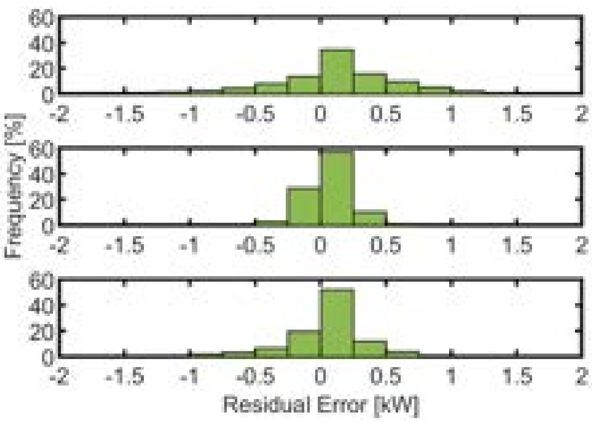

Fig. 9. Residual Error distribution strongly clustered around zero error for conventional (a) retrofit, (b) and NNZE (c) homes based on EnergyPlus synthetic data which was separated validated against experimental data.

variability and the need for an HVAC model at the minute resolution such as the HMLM. An HEM system would benefit from calculations of HVAC system load and PV generation concurrently to ensure balanced operation.

V. Conclusion

The novel hybrid machine learning model (HMLM) proposed in the paper for HVAC systems has been successfully exemplified for two experimental buildings with errors lower than 10%. Key to satisfactory modeling has been the training approach employing k-means clustering. A further advantage on the new HMLM is the requirement of minimal, if any, experimental data. In the examples included in the paper, EnergyPlus models, which have been calibrated against typical one hour data, have been used to derive synthetic data with one minute time resolution, which has been then used for the machine learning algorithm.

The new HMLM black-box, which is employed when the HVAC system is on, is complemented with a thermal-equation white-box simulation for DR type events when the HVAC systems does not draw electric power. The resultant combination is a novel gray-box that is suitable to be used for ultra-fast simulations as a building digital twin model-in-the-loop solution for HVAC electric power and indoor temperature, to quantify cost of operation, economic benefits, and thermal comfort. These simulation capabilities, make the models suitable for studies with industry and utility protocols, such as CTA-2045 and Energy Star for DR events, and for incorporation in co-simulation frameworks for large-scale electric power distribution systems with smart homes and buildings.

Acknowledgment

This paper is based upon work supported by the National Science Foundation (NSF) under Award No. 1936131 and under NSF Graduate Research Fellowship Grant No. 1839289. The support received through a Department of Education (DoEd) GAANN Fellowship is also gratefully acknowledged. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF and DoEd.

References

[1] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, and J. DeGraw, "Grey-box modeling and application for building energy simulations - a critical review," *Renewable and Sustainable Energy Reviews*, vol. 146, p. 111174, 2021.

[2] F. Hafiz, M. Awal, A. R. de Queiroz, and I. Husain, "Real-time stochastic optimization of energy storage management using rolling horizon forecasts for residential pv applications," in *2019 IEEE Industry Applications Society Annual Meeting*. IEEE, 2019, pp. 1–9.

[3] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352710221004356

[4] P. Vongvilasack, S. Premrudeepreechacharn, and K. Ngamsanroaj, "The application of machine learning for the voltage and reactive power control in power distribution network," in *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2022, pp. 1–6.

[5] "Energyplus™, version 00," 9 2017. [Online]. Available: https://www.osti.gov//servlets/purl/1395882

[6] H. Gong, E. S. Jones, A. H. M. Jakaria, A. Huque, A. Renjit, and D. M. Ionel, "Generalized energy storage model-in-the-loop suitable for energy star and cta-2045 control types," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2021, pp. 814–818.

[7] "CTA standard: Modular communications interface for energy management," Consumer Technology Association (CTA), Tech. Rep., 2020.

[8] E. P. R. Institute. (2017) Performance test results: CTA-2045 hvac thermostat. [Online]. Available: https://www.epri.com/research/products/000000003002011747

[9] E. S. Jones, R. E. Alden, H. Gong, A. G. Frye, D. Colliver, and D. M. Ionel, "The effect of high efficiency building technologies and pv generation on the energy profiles for typical us residences," in *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*, 2020, pp. 471–476.

[10] Electric Power Research Institute. (2020) DOE SHINES Residential Demonstration. [Online]. Available: https://dashboards.epri.com/shines-residential/dashboard

[11] E. S. Jones, R. E. Alden, H. Gong, A. A. Hadi, and D. M. Ionel, "Co-simulation of smart grids and homes including ultra-fast hvac models with cta-2045 control and consideration of thermal comfort," in *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2022.