Synthetic Power Consumption Data Generation For Appliance Operation Modes

Abdelkareem Jaradat, Hanan Lutfiyya, Anwar Haque
{ajarada3, hlutfiyy, ahaque32}@uwo.ca
The University Of Western Ontario - Department Of Computer Science
London, Ontario, Canada

***Abstract*—Realistic appliance power consumption data plays a pivotal role in the development of smart home energy management systems and the foundational algorithms for appliance data analysis. However, publicly available datasets are often limited in availability and time-consuming to collect. Consequently, the creation of simulation models for generating synthetic appliance data becomes needed. In this research, a novel approach is designed to simulate power consumption data that is tailored towards appliance operation modes. This model leverages existing public datasets and employs stochastic methods to enhance data variability and consistency. Usage profile characteristics are extracted and used to generate base usage profiles. To synthesize usage profiles while ensuring realism, a probabilistic model is employed and tuning parameters, encompassing components such as white noise, switch-on surges are added to refine the synthetic profiles. The DTW algorithm is then utilized to assess the proximity of the synthetic profiles to the existing ones. Remarkably, our results reveal that the average differences among these profiles can be as low as ten samples, even with a 1Hz sampling frequency.**

***Index Terms*—Appliance Operation Modes; Demand Response (DR); Dynamic Time Warping (DTW); Smart Home Energy Management Systems; SHEMSs; HEMSs; Load Profile Simulation;**

I. Introduction

RECENT research indicates that buildings are responsible for over 40% of the global demand for power consumption and the emissions of greenhouse gases [1]. In Canada, the residential sector's share of total energy usage was 28% in 2006, increasing to 32% in 2020 and continue to rise with the same pattern until 2050 [2]. Particularly, home appliances in a Canadian household account for 14% of total household consumption [3].

The utilization of Smart Home Energy Management Systems (SHEMSs) [4] is a common strategy to reduce residential electric usage. These multi-component systems focus on energy monitoring, analysis, scheduling, and feedback using inputs like electricity tariffs, sensor-collected appliance data, and user preferences. SHEMSs employ Machine Learning and Digital Signal Processing methods to provide user feedback, appliance scheduling, and user information systems [5]. The goal is to enhance user understanding of household usage and promote energy sustainability through approaches such as Demand Response [6].

Within SHEMSs, to develop and validate the aforementioned analytical algorithms and methods, representative datasets are needed. Power Consumption Datasets (PCDs) [1] play a crucial rule in this context. PCDs are datasets containing time-series data corresponding to samples of the instantaneous power consumption for electric loads. Despite the recent efforts in collecting residential PCDs [7], public PCDs availability is still limited [8]. In many cases, researchers use Synthetic PCDs (SyPCDs) [9] to save the installation cost and measurement time [10]. SyPCDs are generated load profiles for household appliances based on either publicly available PCDs (deterministic) or based on mathematical (probabilistic) models [11]. In this work, a novel approach is designed to simulate power consumption data that is tailored towards appliance operation modes. This model is built to extend existing PCDs and aims to simulate household appliances' usage profiles when activated with different Appliance Operation Modes (AOMs) which represent specific settings set by the manufacturer. The main objective of the proposed model is to generate realistic appliance power consumption time series data based on a hybrid model. This model incorporates both deterministic methods that are built on top of a data analysis of publicly available PCDs, and probabilistic methods which adds stochasticity to the model to maximize the realistic aspect of the generated data. The ultimate goal of this model is to generate synthetic usage profile in different AOMs for household appliances.

The rest of the paper is organized as the following: In section II previous related work is presented. In section III the problem formulation is elaborated. Section IV presents the architecture of proposed model. In section V SUPs extraction is discussed. Section VI presents the formal characterization of SUPs. In section VII the process of generating synthetic SUPs is presented. Section VIII evaluates the model. Finally, section IX concludes the paper.

II. Related Work

A probabilistic-empirical residential electricity load model [12] generates 1-minute power usage data for household appliances. This model takes into account both measured data and statistical information, as well as occupant activities. A popular simulator, CREST [13], [14] is also based on active occupancy patterns and occupants' daily activity profiles.

A stochastic approach [15] is used in the generation of high-resolution multi-energy load profiles for residential loads in remote areas. A mathematical framework [16] is developed for simulating household appliances by re-synthesizing the current waveforms, harmonic currents and the phase shifting of the appliances. Similar work [9] uses GUI in Matlab Simulink to simulate household loads.

Generative Adversarial Networks (GANs) have undergone rapid advancements across various domains, including the generation of synthetic PCDs (Power Consumption Datasets). Recent literature, such as TraceGAN [17], ProfileSR-GAN [18],

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

and mREAL-GAN [19], exemplify the exploration of GAN-based methods for creating realistic appliance data.

There are various synthetic datasets available for the residential sector, each serving different purposes. These datasets play a crucial role in the field of residential energy research, enabling researchers to model and analyze various aspects of household energy consumption. One such dataset is the Automated Model Builder for Appliance Loads (AMBAL), which was introduced in the work by Buneeva et al. [20]. AMBAL is designed to simulate appliance loads and create appliance models based on real datasets. Another tool that employs a similar methodology is the SynD dataset [21], which is capable of handling a larger number of appliances. In addition to AMBAL and SynD, there is SmartSim, as described in the research conducted by Chen et al. [22]. SmartSim is a device-accurate home energy load generator that utilizes device energy and device usage models. It simulates household loads through a series of components, including Distribution learning, Event marking, and Trace Generation.

Contribution: While the existing body of literature provides an extensive overview of various methodologies for simulating residential load profiles [23], there exists a research gap in prior research that is specifically dedicated to simulating household appliances within the context of their distinct operation modes. In response, this study contributes by introducing a novel open-source [24] hybrid approach. This approach blends deterministic and probabilistic components, harnessing the combined power of empirical data and statistical models. The result is a robust framework capable of generating appliance usage profiles that encompass the multitude of operation modes encountered in real-world scenarios.

III. Problem Formulation

The purpose of the proposed model is to generate a synthetic dataset of household appliance usage profiles. This section describes the definitions and formulation for the generation process. A Single Use Profile **(SUP)** is used to formally model power consumption of a preprogrammed appliance between the time it is turned on and the time it is switched off. A SUP represents the sequence (time-series) of power consumption values (measured in kW) consumed by an appliance from the moment of turning it on to the moment of turning it off. Typically, home appliances may run in one of several operation modes. An Appliance Operation Mode (AOM) or a program represents a pre-configured setting by the appliance manufacturer that suits the user preferences upon different situations and characterized by its running time and different cycles and states that the appliance passes through. For example, a dishwasher may have three operation modes, a lighter mode for barely used dishes, a medium mode for greasy dishes, and a heavy mode for very greasy dishes. Activating a certain appliance with a certain operation mode consumes electricity differently than other operation modes. According to Makonin et al. [25], the potential saving percentages achieved in load reduction by switching the use of appliances from heavier to lighter AOMs ranges between 25% and 78% for different appliances. For example, if a household switches from using heavy to medium modes in a dryer, 34% of the cost is cut, while if the shifting occurs from heavy towards a light mode, 78% of the consumption is reduced annually.

A daily power consumption sequence, $\mathbf{\Omega}_a^d$, represents the power consumption samples taken in a single day, d, for the appliance, a, such that:

$$\begin{gathered}\mathbf{\Omega}_a^d = \Big\{\omega_n\Big\}_{n=1}^{n_*} = \{\omega_1, \omega_2, \dots, \omega_n, \dots, \omega_{n_*}\} \\ 1 \leq n \leq n_* \end{gathered} \tag{1}$$

where ω_n is the n^{th} instantaneous power sample value measured in (KW), and n_* represents the last sample index in d.

A SUP, ψ^p, with length, θ_{ψ^p}, is defined by the sequence of samples that represent a subsequence of the daily consumption, $\mathbf{\Omega}_a^d$, from the moment of turning the appliance on, n^s, to the moment that it is turned off n^e. This is defined as:

$$\begin{gathered}\psi^p = \{\omega_n\}_{n=n^s}^{n^e} = \{\omega_{n^s}, \omega_{n^s+1}, \dots, \omega_n, \dots, \omega_{n^e-1}, \omega_{n^e}\} \\ \psi^p \subseteq \mathbf{\Omega}_a^d \quad , \quad n^s \leq n \leq n^e \leq n_* \end{gathered} \tag{2}$$

where:

$$\theta_{\psi^p} = n^e - n^s + 1 \tag{3}$$

and for all SUPs, $\psi^p \subseteq \mathbf{\Omega}_a^d$, there is no overlapping among any two SUPs such that the intersection between these SUPs is defined as follows:

$$\begin{gathered}\bigcap_{i,j} \psi_i^{p_j} = \phi \\ \forall i, j \ s.t, \quad 1 \leq i \leq Z_a^d \quad , \quad 1 \leq j \leq P^a \end{gathered} \tag{4}$$

The set of SUPs, $\Psi_a^{d,p}$, of size, $Z_a^{d,p}$, that corresponds to appliance, a , and labeled by the AOM, p, is define as:

$$\begin{gathered}\Psi_a^{d,p} = \{\psi_1^p, \dots, \psi_j^p, \dots, \psi_{Z_a^{d,p}}^p\} \\ p \in \mathbb{P}^a \quad , \quad 1 \leq j \leq Z_a^{d,p} \end{gathered} \tag{5}$$

where $\Psi_a^{d,p}$ contains all the SUPs that run using the same AOM, p. The set of all SUPs, Ψ_a^d, in d is defined as all disjoint subsets $\Psi_a^{d,p}$ corresponding to every AOM $p \in \mathbb{P}^a$. This is defined as:

$$\begin{gathered}\Psi_a^d = \{\psi_1^{p_j}, \dots, \psi_i^{p_k}, \dots, \psi_{Z_a^d}^{p_l}\} \\ \{p_j, p_k, p_l, \dots\} \in \mathbb{P}^a \\ s.t, \quad \bigcap_{p \in P^a} \Psi_a^{d,p} = \phi \end{gathered} \tag{6}$$

where Ψ_a^d is a set of size, Z_a^d, that represents the total size of all AOM subsets such that:

$$\sum_{p \in \mathbb{P}^a} Z_a^{d,p} = Z_a^d \tag{7}$$

The main objective of this model is to generate a set of Synthetic SUPs (SySUPs) that can be used to validate the analytical methods to support Demand Response (DR) [6]. The set of SySUPs, $\ddot{\Psi}_a$, generated by the proposed model is defined as:

$$\ddot{\Psi}_a^p = \mathcal{H}(\Psi_a^p, p, \mathcal{Z}) \tag{8}$$

where $\mathcal{H}$ is the generator function, p is the selected AOM, and $\mathcal{Z}$ is the set of tuning parameters used in the generation process.

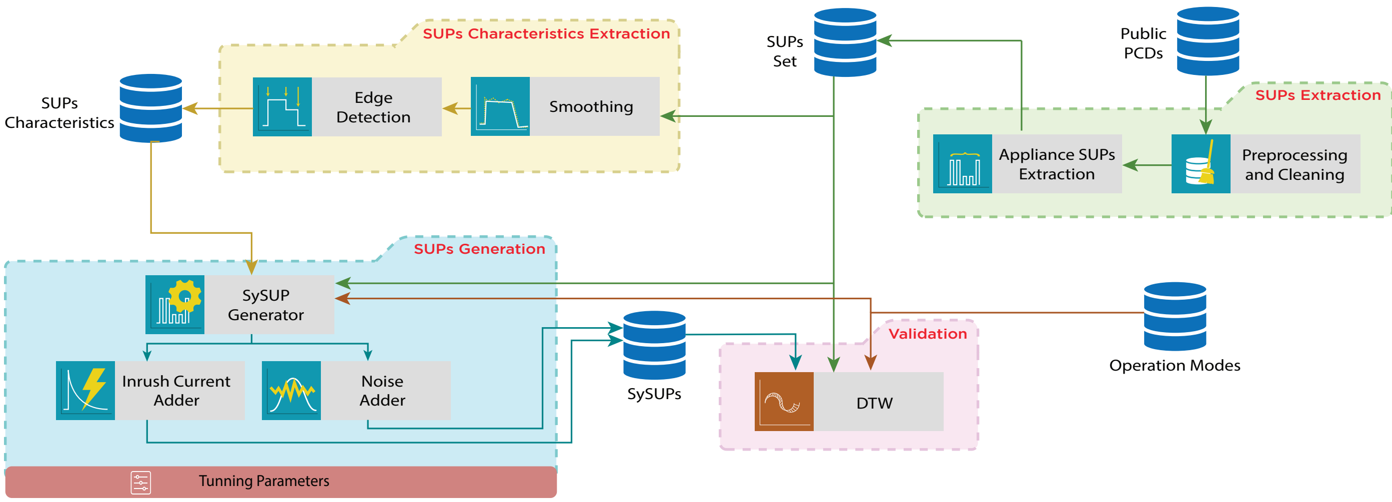

Fig. 1. The architecture used in the proposed approach.

IV. The Architecture

The architecture of this model comprises several integral components. This is depicted in Figure 1. Firstly, the **SUPs Extraction** module undertakes the processing of a publicly available PCD [26], from which it extracts a comprehensive set of Single Use Profiles (SUPs) corresponding to all appliances represented within the PCD. Subsequently, the **SUPs Characteristics Extraction** module engages in a series of processes, aimed at discerning and formalizing the distinctive characteristics inherent in these SUPs. Moving forward, the **SUPs Generation** component takes responsibility for the synthesis of Synthetic SUPs (SySUPs) leveraging the extracted SUPs, their associated characteristics, and the AOMs. This module is thoughtfully crafted, comprising multiple submodules, each contributing to a specific facet of SySUP creation. Lastly, the "Validation" module assumes the task of evaluating the degree of similarity between the resulting SySUPs and the originally extracted SUPs, thus providing a comprehensive assessment of the model's effectiveness.

V. SUPs Extraction

The first step is to isolate the SUPs samples from the rest of the day samples. All SUPs are extracted from the PCDSs [26] on a daily basis per appliance. For a single appliance, the daily consumption sequence $\mathbf{\Omega_a^d}$ contains zero or more SUPs, ψ^p, such that:

$$\begin{gathered}\psi_i^{p_\alpha} = \Big\{\omega_n\Big\}_{n=n^{s_i}}^{n^{e_i}} \quad , \quad \psi_j^{p_\beta} = \Big\{\omega_n\Big\}_{n=n^{s_j}}^{n^{e_j}} \\ i < j \quad \forall \quad \{i,j\} \in \{1,2,\ldots,Z_a^d\} \\ 1 \leq n^{s_i} < n^{e_i} < n^{s_j} < n^{e_j} \leq n_* \end{gathered} \tag{9}$$

where, $\psi_i^{p_\alpha}$, runs with the operation mode, p_α, and starts at $n = n^{s_i}$ and ends at $n = n^{e_i}$. The other SUP, $\psi_i^{p_\beta}$, runs with the operation mode, p_β, and starts at $n = n^{s_j}$ and ends at $n = n^{e_j}$. The activation time of each SUP is before the switch off time as $n^{s_i} < n^{e_i}$, $n^{s_j} < n^{e_j}$ and both SUPs do not overlap throughout the day such that $n^{e_i} < n^{s_j}$. XCorrelation [6] is used to extract the SUPs by sliding a reference SUP over the sequence $\mathbf{\Omega_a^d}$.

A signal can be broken down into multiple individual components. Each component can exhibit a high or low frequency. In this context, a high frequency component with relatively low amplitude is considered noise and needs to be reduced. The moving median smoother is used to reduce the high frequency component. To reduce the high frequency component within SUPs, a transformation function is applied on the SUP sequence, ψ, of length, θ_ψ, to generate the smoothed SUP, $\widehat{\psi}$, of length, $\theta_{\widehat{\psi}}$. The moving median smoother, $\mathcal{M}$, is selected. This transformation is performed as follows:

$$\begin{gathered}\widehat{\psi}(n) = \mathcal{M}(\psi(k)) \\ \forall k \in \Big\{n - \frac{W}{2}, n - \frac{W}{2} + 1, \ldots, n, \ldots, n + \frac{W}{2} - 1, n + \frac{W}{2}\Big\} \\ n \leq \theta_\psi \end{gathered} \tag{10}$$

where, W, is sliding window size that is used by, $\mathcal{M}$.

VI. Extraction of SUP Characteristics

In the context of major appliances, the operational cycle is defined by a series of activations and deactivations of the appliance's internal components. For instance, in the case of a clothes dryer, both the heating element and the spinning motor undergo repetitive on-off cycles throughout its operation. This cyclic operation pattern leads to the emergence of a distinct, recognizable waveform in the corresponding SUPs. These SUPs are often represented as sequences of square-like waves with abrupt transitions (edges) between states, reflecting the switching behavior of internal components. The characteristics of these SUPs, which include the number of states, their distribution, duration, and power levels, play a crucial role in distinguishing the features of SUPs specific to one AOM from those of other AOMs. This distinction is a fundamental aspect of the process of generating SySUPs. The subsequent subsections will elaborate on the steps involved in determining the features of SUPs, with a specific focus on the states that constitute these SUPs.

A. Estimation of state edges

An indicator vector, I, is used to determine the bounds of each state in the SUP, $\widehat{\psi}$. The Median Difference Test (MDT) [27] is used to calculate the values of the indicator vector, I. MDT utilizes a moving window with length, W^I, that slides over the subsequences of $\widehat{\psi}$. The MDT estimates the presence of an edge within $\widehat{\psi}$ by dividing the moving window into two equal length partitions. The median, $\mathcal{M}$, is evaluated for each partition along with the standard deviation, σ, of the entire window. The indicator vector is defined as the following:

$$I^{\widehat{\psi}}(n) = \sqrt{\sigma\left(\widehat{\psi}(\omega^i)\right)\left|\mathcal{M}\left(\widehat{\psi}(\omega^j)\right)^2 - \mathcal{M}\left(\widehat{\psi}(\omega^q)\right)^2\right|} \quad (11)$$
$$\forall\, \omega^i \in k \;,\; \forall\, \omega^j \in k^l \;,\; \forall\, \omega^q \in k^r$$

where $I^{\widehat{\psi}}$ is the indicator vector corresponding to the SUP, $\widehat{\psi}$. $\mathcal{M}(\widehat{\psi}(\omega^j))$ is the median of the SUP samples correspond to the left partition of the window as of $\omega^j \in k^l$, where k^l is the list of samples of the left window. $\mathcal{M}(\widehat{\psi}(\omega^q))$ is the median of the SUP samples that correspond to the right partition of the window as of $\omega^q \in k^r$, where ω^q is the list of samples of the right window. $\sigma(\widehat{\psi}(\omega^i))$ is the standard deviation of the SUP samples of the entire window as of $\omega^i \in k$. The evaluated value of $I^{\widehat{\psi}}(n)$ is proportional to the likelihood of having an edge in $\widehat{\psi}$ at sample index n.

The sequence of thick edges, $\Pi^{\widehat{\psi}}$, of size, $\eta^{\widehat{\psi}}$, that is identified by $I^{\widehat{\psi}}$ in $\widehat{\psi}$, is defined as follows:

$$\Pi^{\widehat{\psi}} = \{\pi_i\}_{i=1}^{\eta^{\widehat{\psi}}} \quad , \quad \eta^{\widehat{\psi}} \le \frac{\theta_{\widehat{\psi}}}{2} \quad (12)$$

where π_i is the i^{th} thick edge that is defined as the pair:

$$\pi_i = (n_i^o, n_i^\iota) \quad , \quad 1 \le n_i^o \le n_i^\iota \le \theta_{\widehat{\psi}} \quad (13)$$

where π_i defines a pair of boundaries, n_i^o as the upper bound, while the lower bound is n_i^ι. The sequence of thick edges, $\Pi^{\widehat{\psi}}$ is obtained by applying a threshold, τ^I, so that the samples below τ^I correspond to the periods outside two thick edges, π_j, π_k. This is defined as the following:

$$\begin{gathered} I^{\widehat{\psi}}(n) \le \tau^I \;,\; \forall\, n \in \left\{n_j^\iota + 1, n_j^\iota + 2, \ldots, n_k^o\right\} \\ \pi_j = (n_j^o, n_j^\iota) \quad , \quad \pi_k = (n_k^o, n_k^\iota) \\ 1 \le n_j^o \le n_j^\iota < n_k^o \le n_k^\iota \le \theta_{\widehat{\psi}} \\ \pi_j, \pi_k \in \Pi^{\widehat{\psi}} \end{gathered} \quad (14)$$

where π_j is the j^{th} thick edge with the lower bound sample index, n_j^o, and the upper bound sample index, n_j^ι.

B. Determining SUP States

A SUP consists of a sequence of states where a state represents the sequence of power values when an internal electrical component within an appliance is activated for a specific period. This sequence follows a relatively constant pattern with slight variation since the internal components consume a steady amount of power.

The sequence of states, Λ, with a size, R, that is associated with $\widehat{\psi}$ is defined as follows:

$$\begin{gathered} \Lambda^{\widehat{\psi}} = \{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_R\} \\ \lambda_i = (e^o, e^\iota, \omega^\lambda) \quad , \quad e^o < e^\iota \\ \lambda_0 = (1, 1, \omega_1^\lambda) \quad , \quad \lambda_R = (\theta_{\widehat{\psi}}, \theta_{\widehat{\psi}}, \omega_{\theta_{\widehat{\psi}}}^\lambda) \end{gathered} \quad (15)$$

where the state, λ_i, is represented by a tuple with three elements. The left exact edge, e^o, the right exact edge, e^ι, and the power value, ω^λ. An Exact Edge is represented by the sample where it is the highest likelihood the abrupt step occurs between two adjacent states. The values of the exact edges are determined through the process of edge thinning. In this process, the exact edge, e, is evaluated from the corresponding thick edge, π. One option for edge thinning is argmax method where the value of e equals the sample index that produces the maximum indicator vector value, $I^{\widehat{\psi}}$. This is defined as:

$$\begin{gathered} e_i^o = argmax\left(I^{\widehat{\psi}}(n)\right) \quad , \quad \forall n \in \pi_i \\ e_i^\iota = argmax\left(I^{\widehat{\psi}}(n)\right) \quad , \quad \forall n \in \pi_{i+1} \\ i < \eta^{\widehat{\psi}} \end{gathered} \quad (16)$$

As both of the left and right exact edges of the state, λ_i are determined, the power value of the state, ω_i^λ, is evaluated as the median, $\mathcal{M}$, of the power values in $\widehat{\psi}$ corresponding to each sample index in the state, λ_i. This is defined as the following:

$$\begin{gathered} \omega_i^\lambda = \mathcal{M}\left(\{\omega_{e_i^o+k}\}\right) \\ \forall\, k \;,\; 0 \le k \le e_i^\iota - e_i^o \end{gathered} \quad (17)$$

where, $\omega_{e_i^o+k}$, is the power values of $\widehat{\psi}$ at the sample index, $e_i^o + k$.

VII. Generating Synthetic SUPs

For a particular day, d, a list of state sequences, $\mathbf{\Lambda}_a^p$, for the set of SUPS, Ψ_a^p, is defined as:

$$\mathbf{\Lambda}_a^p = \left\{\Lambda_i^{\widehat{\psi}_{i,a}^p}\right\}_{i=1}^{Z_a^p} \quad , \quad \forall\, \widehat{\psi}_{i,a}^p \in \widehat{\Psi}_a^p \quad (18)$$

A set of Synthetic SUPs (SySUPs), $\ddot{\Psi}_a^p$, with size, J, is defined as:

$$\ddot{\Psi}_a^p = \left\{\ddot{\psi}_i\right\}_{i=1}^{J} \quad (19)$$

where $\ddot{\psi}_i$ is a SySUP for the appliance, a, with operation mode, p. The value of J represents the synthetic dataset size.

A base SySUP represents a SySUP without adding the effect of any additional tuning parameters. The base SySUP, $\ddot{\psi}$, is defined as follows:

$$\begin{gathered} \ddot{\psi} = \bigcup_{i=1}^{R}\bigcup_{j=1}^{F}\left\{\omega_j^{\lambda_i}\right\} \\ \forall\, \ddot{\psi} \in \ddot{\Psi}_a^p \quad , \quad \forall\, \lambda_i \in \Lambda \quad , \quad \forall\, \Lambda \in \mathbf{\Lambda}_a^p \end{gathered} \quad (20)$$

$$F = e_i^\iota - e_i^o \quad (21)$$

where ω^{λ_i} is the power value of the state, $\lambda_i \in \Lambda$. R, is the size of the states sequence, Λ. And, F, is the length of the state, λ_i. The following subsections discuss two probabilistic components that is added to the base SySUP.

A. The white noise component

Since the states of a SUP are not completely flat, an added noise to the SySUP is required to simulate more realistic SUP states. The noise coefficient, ξ, is defined as follows:

$$\xi_j = \mathcal{N}\left(\mu^{\xi}, \sigma^{\xi}\right) \tag{22}$$

where ξ_j is the added noise to the j^{th} sample in the base SySUP, $\ddot{\psi}$. This noise is selected based on a normal distribution function, $\mathcal{N}$, with a mean of μ^{ξ} and standard deviation σ^{ξ}. The set of SySUPs with the added noise, ${}^{\xi}\ddot{\Psi}^{p}_{a}$, is defined as:

$${}^{\xi}\ddot{\Psi}^{p}_{a} = \left\{ {}^{\xi}\ddot{\psi}_i \right\}_{i=1}^{J} \tag{23}$$

where the resulting SySUP with the added noise, ${}^{\xi}\ddot{\psi}$, is defined as:

$${}^{\xi}\ddot{\psi} = \bigcup_{i=1}^{R} \bigcup_{j=1}^{F} \left\{ \omega_j^{\lambda_i} + \xi_j \right\} \quad \forall\ \lambda_i \in \Lambda \tag{24}$$

B. The switch-on surge component

Inrush current, or the switch-on surge (SOS) is the maximum instantaneous input current consumed by electrical transformers within an electrical device when first switched on [28]. This current appears within the first few samples of high-power states when a major component within the appliance is triggered. The SOS component appears at the beginning of a state as a sharp spike that eventually starts decaying in its amplitude. The set of SySUPs with the added SOS, ${}^{\vartheta}\ddot{\Psi}^{p}_{a}$, is defined as:

$${}^{\vartheta}\ddot{\Psi}^{p}_{a} = \left\{ {}^{\vartheta}\ddot{\psi}_i \right\}_{i=1}^{J} \tag{25}$$

where the resulting SySUP with the added SOS, ${}^{\vartheta}\ddot{\psi}$, is defined as:

$${}^{\vartheta}\ddot{\psi} = \bigcup_{i=1}^{R} \bigcup_{j=1}^{F} \left\{ \omega^{\lambda_i} + \xi_j + \frac{\vartheta_j}{j} \right\} \quad , {}^{\vartheta}\ddot{\psi} \in {}^{\vartheta}\ddot{\Psi}^{p}_{a} \quad \forall\ \lambda_i \in \Lambda \tag{26}$$

where ϑ_j is the SOS coefficient that follows a normal direction function as:

$$\vartheta = \mathcal{N}\left(\mu^{\vartheta}, \sigma^{\vartheta}\right) \tag{27}$$

The term $\frac{\vartheta_j}{j}$ represents the positive side of a hyperbola in which this modeling of the SOS component simulates the behavior of SOS current at each state.

VIII. Evaluation

To evaluate the impact of the tunning parameters explained in the previous section on the SySUP with respect to the SUP, an evaluation metric, $\bar{\delta}$, is defined as the following:

$$\bar{\delta}\left({}^{\xi}\ddot{\psi}^{p}_{a}, \Psi^{p}_{a}\right) = \frac{1}{J}\sum_{j=1}^{J}\left[\frac{1}{Z_a^{d,p}}\sum_{i=1}^{Z_a^{d,p}}\frac{DTW\left({}^{\xi}\ddot{\psi}^{p}_{j}, \psi^{p}_{i}\right)}{\theta_{\psi^p_i} + \theta_{\ddot{\psi}^p_j}}\right] \tag{28}$$

where $\bar{\delta}$ represents the average DTW distance [29] between a SySUP, ${}^{\xi}\ddot{\psi}^{p}_{a} \in {}^{\xi}\ddot{\Psi}^{p}_{a}$, and every SUP, $\psi^{p}_{a} \in \Psi^{p}_{a}$, for the appliance, a, using the operation mode, p. The other evaluation metric, $\bar{\kappa}$, is defined as:

$$\bar{\kappa}\left({}^{\xi}\ddot{\psi}^{p}_{a}, \Psi^{p}_{a}\right) = \frac{1}{J}\sum_{j=1}^{J}\sigma\left(\left\{\frac{1}{Z_a^{d,p}}\sum_{i=1}^{Z_a^{d,p}}\frac{DTW\left({}^{\xi}\ddot{\psi}^{p}_{j}, \psi^{p}_{i}\right)}{\theta_{\psi^p_i} + \theta_{\ddot{\psi}^p_j}}\right\}_{j=1}^{J}\right) \tag{29}$$

where $\bar{\kappa}$ represents the pooled standard deviations of DTW distance between each SySUP, ${}^{\xi}\ddot{\psi}^{p}_{a} \in {}^{\xi}\ddot{\Psi}^{p}_{a}$, and every SUP, $\psi^{p}_{a} \in \Psi^{p}_{a}$, for the appliance, a, using the operation mode, p.

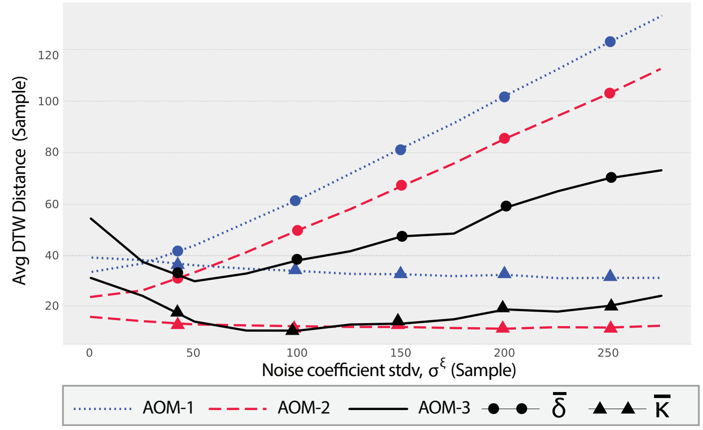

Fig. 2. The impact of changing the noise coefficient, ξ, on the valued of the distance mean, $\bar{\delta}$, for a dryer.

A. Evaluating the white noise component

In Figure 2, the impact of changing the noise coefficient, ξ, on the value of the distance mean, $\bar{\delta}$, is demonstrated for a dryer. Figure 2 shows the values of σ^{ξ} in Eq 22 in the range $\sigma^{\xi} \in [1, 300]$ samples while distribution mean is $\mu^{\xi} = 0$. The plot shows 3 operation modes for the dryer each of which has $\bar{\delta}, \bar{\kappa}$ plots.

The metric,$\bar{\delta}$, reflects the similarity between SUPs and SySUPs. Generally, $\bar{\delta}$ plots starts flat or decaying and then it shows a continuous increase. For example, in Figure 2, the AOM-2 plot is decreasing when $1 \leq \sigma^{\xi} \leq 50$ which means that the added noise contributes in increasing the similarity among SUPs and SySUPs. It then starts increasing for the rest of $\bar{\delta}$ values which means higher noise values result in a decrease in similarity.

On the other hand, the metric, $\bar{\kappa}$, reflects the consistency of the other metric values, $\bar{\delta}$. Lower values of $\bar{\kappa}$ reflect consistent distances among SUPs and SySUPs. However, higher values of $\bar{\kappa}$ correspond to high variation in $\bar{\delta}$ which means the distances have large differences within the same experiment. As an example, in Figure 2, the $\bar{\kappa}$ plot for AOM-3 has less fluctuation which means high consistency in $\bar{\delta}$.

B. Evaluating the switch-on surge component

To evaluate the impact of the SOS coefficient, ϑ, Eq 28 and Eq 29 are used. The distance metric, $\bar{\delta}\left({}^{\vartheta}\ddot{\psi}^{p}_{a}, \Psi^{p}_{a}\right)$, evaluates the impact of the SOS coefficient, ϑ, on the SySUP, ${}^{\vartheta}\ddot{\psi}^{p}_{a}$,

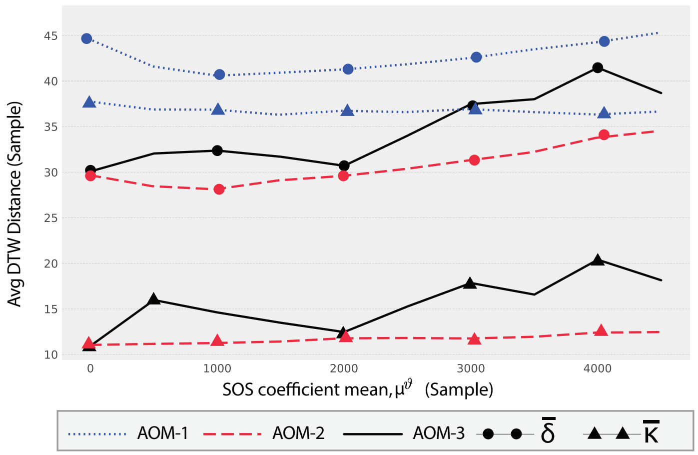

Fig. 3. The impact of changing the SOS coefficient, ϑ, on the valued of the distance mean, $\bar{\delta}$, for a dryer.

with respect to the SUPs, Ψ_a^p, while $\bar{\kappa}\left({}^{\vartheta}\ddot{\psi}_a^p, \Psi_a^p\right)$ reflects the consistency of distances, $\bar{\delta}$. This is illustrated in Figure 3 which shows the values of μ^{ϑ} in Eq 27 in the range $\mu^{\vartheta} \in [1, 5000]$ samples while distribution mean is $\sigma^{\vartheta} = 100$. The plot shows 3 operation modes for the dryer each of which has $\bar{\delta}, \bar{\kappa}$ plots. Both metrics, $\bar{\delta}, \bar{\kappa}$ follow the same rhythm in Figure 2 where $\bar{\delta}$ starts with a relatively higher value, then starts to decay and then begin rising again. The $\bar{\kappa}$ metric shows a steady response for the majority of μ^{ϑ} values.

IX. Conclusion

The expanding landscape of analytical methods for processing power consumption data has encouraged researchers to explore the use of simulation techniques, in order to extend the availability of public datasets. These datasets serve as a valuable testing ground for validating the efficacy of analytical algorithms. In this study, a tailored approach to simulating power consumption data, specifically focusing on appliance operation modes is proposed. This model leverages a combination of deterministic and probabilistic formal models to mimic household appliance usage patterns across various operation modes. By employing the Dynamic Time Warping (DTW) algorithm to assess the model's performance, we demonstrated a remarkable similarity between the original and the synthetic usage profiles with an average difference of just 10 samples at a 1Hz sampling rate. Next steps may involve expanding the number of tunning parameters used in the simulation.

References

[1] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Building power consumption datasets: Survey, taxonomy and future directions," *Energy and Buildings*, p. 110404, 2020.

[2] "Canada energy regulator, canada's energy future," 2021. [Online; accessed 29-May-2023]. https://bit.ly/3d8vIqH.

[3] N. R. Canada, "Natural resources canada, appliances for residential use," 2021. [Online; accessed 29-May-2023]. https://bit.ly/3omEb0d.

[4] A. AlHammadi, A. AlZaabi, B. AlMarzooqi, S. AlNeyadi, Z. AlHashmi, and M. Shatnawi, "Survey of iot-based smart home approaches," in *2019 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1–6, IEEE, 2019.

[5] A. Jaradat, H. Lutfiyya, and A. Haque, "Smart home energy visualizer: A fusion of data analytics and information visualization," *IEEE Canadian Journal of Electrical and Computer Engineering*, 2022.

[6] A. Jaradat, H. Lutfiyya, and A. Haque, "Demand response for residential uses: A data analytics approach," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pp. 1–6, IEEE, 2020.

[7] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Building power consumption datasets: Survey, taxonomy and future directions," *Energy and Buildings*, vol. 227, p. 110404, nov 2020.

[8] G. Flett and N. Kelly, "Modelling of individual domestic occupancy and energy demand behaviours using existing datasets and probabilistic modelling methods," *Energy and Buildings*, vol. 252, p. 111373, 2021.

[9] J. M. G. Lopez, E. Pouresmaeil, C. A. Canizares, K. Bhattacharya, A. Mosaddegh, and B. V. Solanki, "Smart residential load simulator for energy management in smart grids," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1443–1452, 2018.

[10] N. Pflugradt and U. Muntwyler, "Synthesizing residential load profiles using behavior simulation," *Energy Procedia*, vol. 122, pp. 655–660, 2017.

[11] M. Kabirifar, N. Pourghaderi, A. Rajaei, M. Moeini-Aghtaie, and A. Safdarian, "Deterministic and probabilistic models for energy management in distribution systems," in *Handbook of Optimization in Electric Power Distribution Systems*, pp. 343–382, Springer, 2020.

[12] A. Marszal-Pomianowska, P. Heiselberg, and O. K. Larsen, "Household electricity demand profiles–a high-resolution load model to facilitate modelling of energy flexible buildings," *Energy*, vol. 103, pp. 487–501, 2016.

[13] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy and Buildings*, vol. 42, no. 10, pp. 1878–1887, 2010.

[14] I. Richardson and M. Thomson, "Domestic electricity demand model - simulation example," Jan 2010.

[15] F. Lombardi, S. Balderrama, S. Quoilin, and E. Colombo, "Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model," *Energy*, vol. 177, pp. 433–444, 2019.

[16] C. B. Nilsen, B. Hoff, and T. Østrem, "Framework for modeling and simulation of household appliances," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 3472–3476, IEEE, 2018.

[17] A. Harell, R. Jones, S. Makonin, and I. V. Bajić, "Tracegan: synthesizing appliance power signatures using generative adversarial networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4553–4563, 2021.

[18] L. Song, Y. Li, and N. Lu, "Profilesr-gan: A gan based super-resolution method for generating high-resolution load profiles," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3278–3289, 2022.

[19] E. Sanderson, A. Fragaki, J. Simo, and B. J. Matuszewski, "mreal-gan: Generating multiple residential electrical appliance load profiles with inter-dependencies using a generative adversarial network," *arXiv preprint arXiv:2112.06656*, 2021.

[20] N. Buneeva and A. Reinhardt, "Ambal: Realistic load signature generation for load disaggregation performance evaluation," in *2017 ieee international conference on smart grid communications (smartgridcomm)*, pp. 443–448, IEEE, 2017.

[21] C. Klemenjak, C. Kovatsch, M. Herold, and W. Elmenreich, "A synthetic energy dataset for non-intrusive load monitoring in households," *Scientific data*, vol. 7, no. 1, p. 108, 2020.

[22] D. Chen, D. Irwin, and P. Shenoy, "Smartsim: A device-accurate smart home simulator for energy analytics," in *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 686–692, IEEE, 2016.

[23] G. Happle, J. A. Fonseca, and A. Schlueter, "A review on occupant behavior in urban building energy models," *Energy and Buildings*, vol. 174, pp. 276–292, 2018.

[24] A. Jaradat, "Github repo for hydrosafe," 2021. [Online; accessed 29-May-2023]. github.com/abedjar/HYDROSAFE.

[25] S. Makonin, Z. J. Wang, and C. Tumpach, "Rae: The rainforest automation energy dataset for smart grid meter data analysis," *data*, vol. 3, no. 1, p. 8, 2018.

[26] S. Makonin, Z. J. Wang, and C. Tumpach, "RAE: The Rainforest Automation Energy Dataset for Smart Grid Meter Data Analysis," pp. 1–9, 2017.

[27] R. Fried, "On the robust detection of edges in time series filtering," *Computational Statistics & Data Analysis*, vol. 52, no. 2, pp. 1063–1074, 2007.

[28] K. J. Patel, "Effects of transformer inrush current," 2013.

[29] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series.," in *KDD workshop*, vol. 10, pp. 359–370, Seattle, WA, USA:, 1994.