Energy Management of Buildings with Energy Storage and Solar Photovoltaic: A Diversity in Experience Approach for Deep Reinforcement Learning Agents

Akhtar Hussain[a,*], Petr Musilek[a,b]

[a]*Department of Electrical and Computer Engineering, University of Alberta, Edmonton, T6G 2G2, AB, Canada*
[b]*Department of Applied Cybernetics, University of Hradec Králové, 500 03, Hradec Králové, Czech Republic*

Abstract

Deep reinforcement learning (DRL) is a suitable approach to handle uncertainty in managing the energy consumption of buildings with energy storage systems. Conventionally, DRL agents are trained by randomly selecting samples from a data set, which can result in overexposure to some data categories and under/no exposure to other data categories. Thus, the trained model may be biased toward some data groups and underperform (provide suboptimal results) for data groups to which it was less exposed. To address this issue, diversity in experience-based DRL agent training framework is proposed in this study. This approach ensures the exposure of agents to all types of data. The proposed framework is implemented in two steps. In the first step, raw data are grouped into different clusters using the K-means clustering method. The clustered data is then arranged by stacking the data of one cluster on top of another. In the second step, a selection algorithm is proposed to select data from each cluster to train the DRL agent. The frequency of selection from each cluster is in proportion to the number of data points in that cluster and therefore named the *proportional selection method*. To analyze the performance of the proposed approach and compare the results with the conventional random selection method, two indices are proposed in this study: the flatness index and the divergence index. The model is trained using different data sets (1-year, 3-year, and 5-year) and also with the inclusion of solar photovoltaics. The simulation results confirmed the superior performance of the proposed approach to flatten the building's load curve by optimally operating the energy storage system.

Keywords: Battery energy storage, building demand management, deep reinforcement learning, diversity in experience, energy management.

1. Introduction

Buildings are one of the main energy consumption hubs and account for about 30% of total energy consumption [1]. Building energy demand has increased by about 4% in 2021 compared to the previous year, which is the largest increase in the last 10 years [2]. About

*Corresponding author
Email address: akhtar3@ualberta.ca (Akhtar Hussain)

31% of the building energy demand is fulfilled directly by electricity. It has been reported in the Electricity Market Report 2023 [3], that the global electricity demand increased by about 2% in 2022 as compared to the previous year, despite the global energy crisis. In the same report it is predicted that global energy demand will increase by about 3.2% per year from 2024 onward, which is much higher than the pre-pandemic rate of 2.4%. The increase in electricity demand is expected to be due to transportation electrification [4], electrification of heating systems [5], and the production of alternative fuels such as hydrogen on an industrial scale [6]. It has been reported in [7] that replacement of furnaces with reversible heat pumps will increase the electricity demand (in winter) by up to 36%.

Excessive energy consumption in buildings makes them a major source of carbon emissions [8] that can be mitigated by integrating renewable energy sources (RES). The integration of RES, especially rooftop solar photovoltaic (PV), has gained momentum recently [9]. However, isolated PV use can result in higher system ramp rates [10] due to reduced PV power and increased residential load in the evenings. In addition, the peak PV power coincides with the lower load intervals in residential buildings. Therefore, the integration of PV alone is not sufficient to mitigate the evening peak loads. To address this issue, the integration of battery energy storage systems (BESSs) is considered in several studies, as discussed below.

An analysis of the economic potential of BESSs in buildings [11] concluded that operation cost can be reduced by up to 5.3% with the integration of BESS alone. Different storage options to manage the electrical load of buildings with PVs are analyzed in [12] and identifying lithium-ion batteries as one of the viable storage options. Different design aspects of BESSs such as home energy storage and community energy storage, to manage building load, are analyzed in [13]. The profitability of integrating BESS with PV in buildings is analyzed in [14]. The authors note that BESS with PV is more profitable compared to PV alone if all incentives are taken into account. However, integration of BESS alone or BESS with PV in buildings requires a building energy management system (BEMS) to schedule the BESS operation and minimize the electricity bills of the building. Several studies on managing the load of buildings using BEMS are discussed as follows.

The existing literature on BEMS can be broadly divided into two categories, model-based operation techniques [15] and reinforcement/deep reinforcement learning (DRL)-based operation techniques [16]. In the former, a mathematical model is developed with the goal of minimizing/maximizing a particular objective function under some constraints. The uncertainty of the model can be described using robust [17] or stochastic optimization methods [18]. In the latter, a reward function is devised to train a DRL agent using diversified data, and the trained model is later used for real-time operation [19]. Different model-based BEMSs are proposed in the literature. For example, a Lagrangian multiplier-based model is developed to maximize revenue with BESS constraints under the smart grid paradigm in [19]. Similarly, a multiobjective optimization problem is formulated in [20] for reducing the energy bills of a residential building while reducing the peak load of the system. The number of constraints in model-based approaches can be overwhelming with increasing number of appliances/equipment. Therefore, a method to reduce the number of constraints is proposed in [21] without significantly affecting the optimal results.

Similarly, several studies have used reinforcement learning and DRL for managing building loads, including BESS and PV [22]. For example, a Q learning model is developed in [23] to control BESS energy and reduce electricity bills in a residential building. Similarly, fuzzy

logic is used in the reward function of a Q-learning algorithm in [24] to schedule the home appliance of a smart home, including electric vehicles and BESS. However, reinforcement learning algorithms are unable to handle uncertainties. Therefore, several researchers have proposed DRL-based energy management schemes. For example, the deep deterministic policy gradient is used by [25] and [26] for realizing BEMSs. The method proposed in [25] does not require the dynamic model of the system, and a multi-agent system with prioritized experience replay is considered in [26]. Similarly, soft actor-critic (SAC) is used in [27] and [28] to manage building load. Both cluster energy cost and peak load are considered in [27] and the average peak load and average-to-peak load ratio are considered in [28]. Application of DRL for different types of buildings such as residential, commercial, offices, data centers, and educational are discussed in [29].

However, several challenges need to be addressed for the adoption of RL/DRL for controlling real buildings. For instance, three primary barriers to the adoption of RL in real buildings have been identified in [30]: the substantial amount of data required, control security, and the inability to leverage transfer learning. Similarly, [31] outlines ten major challenges in the real-world implementation of RL for building energy management, categorized as learning, infrastructure, cost/benefit, and safety/security challenges. A computer science perspective on the application of RL to buildings is presented in [32], which divides buildings into single and clustered units while considering single and multi-agent aspects. Challenges in the application of RL for grid-integrated buildings are discussed in [33], where a total of nine significant challenges are identified, including poorly defined objective functions and learning from limited samples. It is evident that most studies have recognized challenges related to the efficient utilization of available data, such as data requirements for training and learning from limited samples.

Therefore, efficient exploration is considered a challenging issue in DRL problems since the agent can only learn from the experiences acquired from the environment [34], [35]. It becomes especially challenging when the problem has large state space and/or sparse rewards [36]. To address this issue, the authors add a regularization based on a distance measure to the loss function to enhance the agent's exploration behavior [36]. Local and global goals are defined for agents to enhance the diversity of the population in [34]. Two different generators are proposed to enhance the exploration potential of DRL agents in [37]. One generator is more involved with exploration and the other with exploitation while having the same architecture. In all these studies, the performance of the DRL agent has improved. In most cases, the improvement is due to better exploration.

It is important to note that [34], [36], and [37] propose different approaches to modify the internal architecture of the model. To achieve this, they introduce additional hyperparameters, which can lead to complexity during the training process and necessitate tuning for each specific problem. Moreover, the policy embeddings suggested in [36] rely on heuristics and might not be easily generalized to new environments, as pointed out in [34]. Notably, the technique outlined in [36] is better suited for games where rewards are sparse and misleading. Similarly, the methodologies introduced in [34] and [37] draw inspiration from novelty search methods, which concentrate on exploring unique solution spaces. Consequently, these approaches might prioritize solutions that are innovative but not necessarily effective. Furthermore, none of these methods have taken into account a balanced mixture of experiences. This aspect holds particular relevance for building energy management, given that the en-

ergy profiles of buildings do not undergo significant changes on a daily basis. Therefore, a tailored approach is essential to ensure diversity in the learning process of DRL agents aimed at handling building loads.

In addition, the energy management of buildings is a large state-space problem since it needs data for several years and data changes during each time interval. This is because the energy consumption pattern of buildings changes significantly across different seasons of the year. Similarly, the demand profile could be different for working days and holidays, and during different hours of the day. Therefore, the DRL agent should be able to optimally manage the load under these diverse consumption patterns. However, conventionally, the DRL agent is trained by selecting random profiles from the data set [25], [28]. This may lead to over-exposure of the agent to one category of load profiles and under-exposure/no exposure to another category, thus resulting in suboptimal solutions for the latter categories. With conventional random selection, the agent may revisit previously explored states and thus limit the coverage of the entire state space [38]. In addition, it may result in the instability of the model due to frequent changes in the reward. Since the DRL agent learns by interacting with the environment, exposure to diverse environments is necessary to ensure its optimal operation [39]. Particularly, it is necessary to ensure the diversity in experience of the DRL agent to different possible state spaces/environments (load profiles in this case). However, ensuring diversity in experience of the agent is challenging, as the data need to be organized into different groups (clusters) prior to passing it to the DRL agent. In addition, separate algorithms are required for selecting data, during each episode, from the data set to ensure the inclusion of all categories of data.

To address these issues, a diversity in experience-based approach is proposed to train DRL agents to manage building loads with BESS and PV. Unlike existing studies on the diversity of DRL agents [34], [36], [37], where the internal architecture of the model is modified, the proposed method is based on clustering of the input data set. Therefore, the proposed method can ensure the diversity of experience without altering the internal architecture of the DRL model. In the proposed method, first, the daily load profiles of the building are clustered into different groups. The K-means clustering method is used to divide the daily load profiles into K clusters and the optimal value of K for this analysis turns out to be 6. Then, the data corresponding to each cluster are extracted and arranged (stacked on top of each other). A selection algorithm is devised to select the data from each group in proportion to its size (the number of samples it contains) and is called the *proportional selection method*. These steps can be integrated into the data preprocessing phase. The DRL agent is trained using the selected data, which ensures the exposure of the agent to different data clusters. Two indices, named the flatness index and the divergence index, are formulated to analyze and compare the performance of the proposed approach with that of the conventional random selection approach. The flatness index measures the flatness of the load curve by comparing the peak and valley loads. The divergence index measures the difference between each data point and the average daily load. Simulations are carried out using different data sets, including 1-year, 3-year, and 5-year data, and also with the inclusion of PV in the building. Finally, detailed analyses are conducted for selected days of the year by taking one sample from each cluster. For the sake of simplicity, the proposed approach based on the proportional selection method is named the proportional selection method (PSM), and the conventional random selection method is named the random selection method (RSM) throughout the remainder

of the paper.

The remainder of the paper is organized as follows. The introduction section is followed by the background information required to understand the proposed energy management approach based on diversity in experience, discussed in Section 2. SAC-based agent modeling and training are discussed in Section 3. The performance of the PSM and RSM are compared in Section 4 in terms of convergence and stability. Detailed analysis using different datasets is carried out in Section 5 which is followed by conclusions, Section 6.

2. Diversity in Experience

Diversity in the experience of a DRL agent is a measure of how well the agent has been trained with exposure to different available state-space groups/clusters of the environment. In DRL, an agent reaches a behavioral policy to maximize the reward by interacting with the environment [39]. Random selection of samples from the data, during training, may result in biased training of the agent towards a particular data category, for example, summer weekdays or winter holidays in our case. To address this issue, preprocessing of data is required to group the data into different clusters. Then, a selection algorithm can be used to select samples from each cluster, for training, considering the size of each cluster. Details of the proposed model are presented in the next section. In the subsequent subsections, an overview of the system under consideration and the clustering of data into different representative groups is discussed.

2.1. System Configuration

The system considered in this study is a residential building; it could be a single detached home or an apartment with local energy storage and may also have solar PV. An overview of the system configuration is shown in Fig. 1. It can be seen that the building contains a building load, solar PV, and a BESS unit. For the sake of simplicity, the building appliances are not shown separately, since the focus of this study is on training the DRL agent to optimize the operation of the BESS. The building is connected to the utility grid and can buy power from the grid when the local energy is not sufficient to meet the energy demand. A BEMS is responsible for managing the operation of the building. The DRL agent is a part of the BEMS and the steps involved in the training of the DRL agent, under the PSM, are as follows.

First, the daily energy demand profiles of the building are gathered for a sufficient duration to include diverse load patterns across different seasons of the year and days of the week. The data are then clustered into several distinct groups to ensure the representation of each group. Details about the clustering analysis are presented in the following section. Then a selection algorithm (explained in Section 3) is used to select samples from each cluster for the training of the DRL agent. Finally, a trained model, with the experience of all representative data groups, is obtained and can be used for optimal building operation with BESS and PV.

2.2. Clustering of Daily Energy Demand Profiles

The daily energy demand profiles of the building are first divided into different groups to train the DRL agent. K-means clustering is the most widely used technique for unsupervised clustering. In K-means clustering, an n-dimensional data set is divided into K clusters

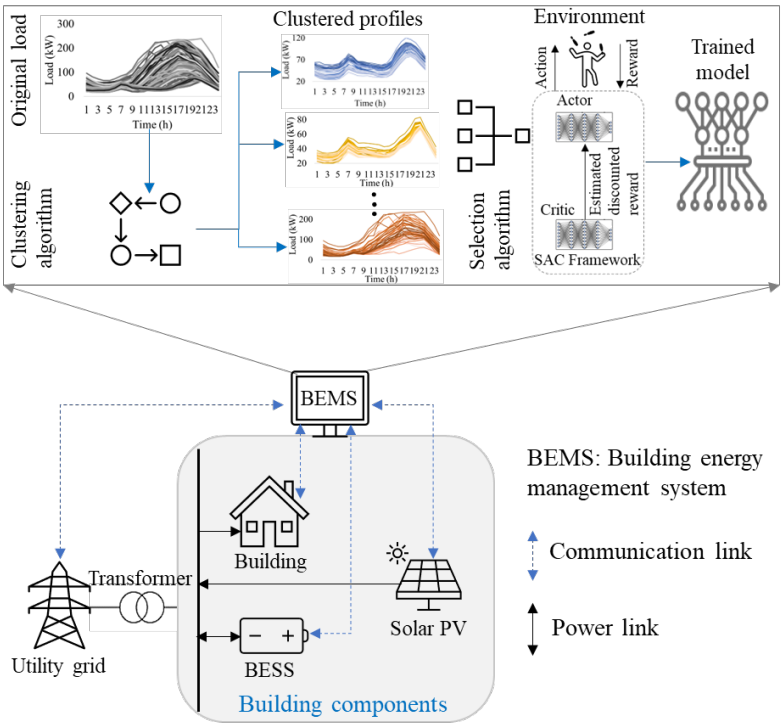

Figure 1: Overview of the system configuration and components of the proposed method.

with the objective of minimizing the sum of squares within each partition [40]. Euclidean distance is generally used as a measure to determine the similarity between two objects [41]. Determination of the optimal number of clusters is required for any dataset, and then the number of data points in each cluster is determined. An overview of the procedure for determining the optimal number of clusters for building load profiles using the K-means method is shown in Algorithm 1.

Algorithm 1: Optimal number of clusters using the K-means method.

Data: Daily energy consumption profiles of the building for 6 years
Preprocess the data and normalize using the min-max scalar
Initialize the maximum number of clusters (C)
for *j=1:C* **do**
    Randomly determine cluster centroids (c_j)
    while *Centroid position is changed* **do**
        Compute the sum of the squared distance between all data points (1)
        Reassign each element to the cluster to minimize the distance
    end
    Record the distortion for each run
    Record the computation time for each run
end
Result: Optimal number of clusters using the elbow method

First, the daily load profiles of the building for a specific period (6 years in this study) are taken as input. The data are then normalized using the min-max function to eliminate the domination of higher-magnitude profiles. Then, the maximum number of clusters to be

tested (C) is selected. An outer loop is used to vary the number of clusters from 1 to C. For each value of j, a cluster centroid c_j is defined and an inner loop is used to assign the data points to each cluster. The assignment is carried out by computing the Euclidean distance (J) using the following equation

$$J = \sum_{j=1}^{C} \sum_{i=1}^{N} ||x_i^{(j)} - c_j||^2, \tag{1}$$

where $x_i^{(j)}$ is the data point i selected for cluster j and c_j is the j^{th} centroid. Each data point is assigned to its nearest centroid based on the distance J. The centroids are then recomputed by taking the mean of all data points within the cluster. This process is repeated until no change is observed in the data points' position in clusters. Finally, the distortion and computation time of each cluster are recorded, and the optimal number of clusters is determined using the elbow method. The results of the optimal number of clusters are shown in the simulations section.

2.3. Proportional Selection Method

The daily load profiles of the building are grouped into N clusters, as discussed in the previous section. Next, a sample selection algorithm is required to select data samples from each cluster during the training process. Details about the selection algorithm are provided in the next section. Here, the basic working principle of the PSM is compared with that of the RSM. An overview of both selection methods is provided in Fig. 2. In this illustration, $N = 4$, and each cluster is represented by a different color. It is also assumed that each set includes 4 episodes, where an episode refers to the actions and states an agent undergoes from the start to the end state.

It can be observed that in the case of the conventional method (RSM), the four samples (for each episode) are randomly selected from the data set that has D samples. Therefore, different clusters are dominant during different episodes. For example, profile type 1 is dominant (2 out of 4 samples) in set 1, and profile type 2 is dominant in set 2 (3 out of 4 samples). The conventional method is not only unable to ensure the diversity in experience of the agent, but would also likely result in instability of the model due to fluctuating average rewards in each set.

In the case of the PSM, the number of samples in each cluster is counted first, that is, $D1, D2, D3$, etc. The samples from each group are then selected considering the ratio of the number of samples in that cluster to the total number of samples D. For example, in this illustration, 25% of the samples belong to cluster 1, 50% to cluster 2, and the remaining 25% to cluster 3. Therefore, in each set, two samples are selected from cluster 2 and one each from clusters 1 and 3, i.e., *proportional selection method*. The PSM ensures that the agent is exposed to all types of datasets by explicitly including samples from each cluster. In addition, it will likely increase the stability and expedite the convergence speed of the average reward due to similar profiles in each training set.

2.4. System Model

In this section, a mathematical model is developed for the energy storage system and the constraints of the network power balance are introduced. Note that these equations cannot

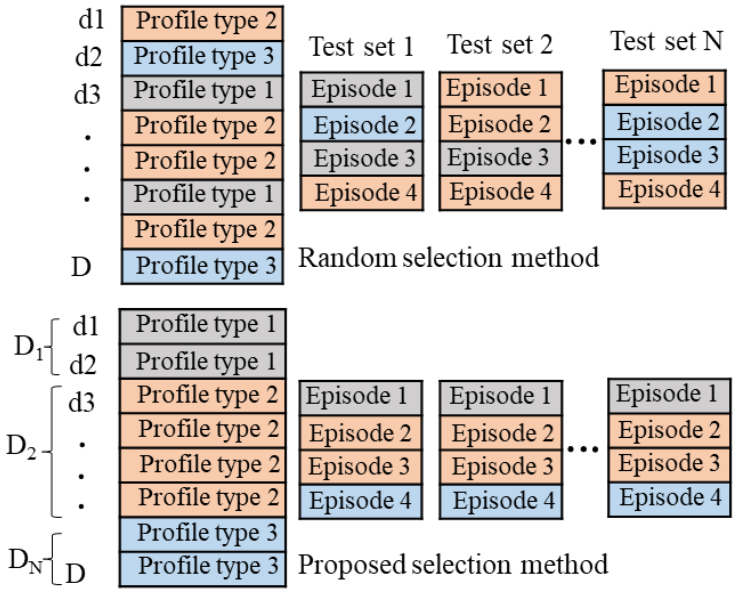

Figure 2: Overview of sample selection methods for training of DRL agent.

be used directly, as they are intended for the training of DRL agents. Their equivalent counterparts will be devised using if-else conditions. The upper ($SoC^{\max}$) and lower ($SoC^{\min}$) state-of-charge (SoC) limits are introduced to avoid overcharging and deep discharging of the battery. Therefore, the amount of power charged to the battery (P_{τ}^{char}) and the amount of power discharged from the battery (P_{τ}^{disc}) during any time interval t should not violate those limits. The upper limit constraint is realized by using the following equation

$$E^{\text{ini}} + \sum_{\tau \leq t} \eta^{+} \times P_{\tau}^{\text{char}} \times \Delta t - \frac{P_{\tau}^{\text{disc}} \times \Delta t}{\eta^{-}} \leq \frac{B^{\text{cap}} \times SoC^{\max}}{100}, \tag{2}$$

where E^{ini} is the initial amount of energy in the BESS, and B^{cap} is the BESS capacity in kWh. The parameters η^{+} and η^{-} represent the charging and discharging efficiencies of the BESS, respectively. Finally, Δt represents the length of the time period, in hours, during which the BESS is charged or discharged. Similarly, the lower limit constraint is realized using the following equation

$$E^{\text{ini}} + \sum_{\tau \leq t} \eta^{+} \times P_{\tau}^{\text{char}} \times \Delta t - \frac{P_{\tau}^{\text{disc}} \times \Delta t}{\eta^{-}} \geq \frac{B^{\text{cap}} \times SoC^{\min}}{100}. \tag{3}$$

The amount of charging and discharging power in BESS is also constrained by the charging and discharging rates of the converter. The charging (P^{Crate}) and discharging (P^{Drate}) rate constraints are enforced using the following equation

$$0 \leq P_{t}^{\text{char}} \leq P^{\text{Crate}}, 0 \leq P_{t}^{\text{disc}} \leq P^{\text{Drate}}. \tag{4}$$

Finally, the building can fulfil its energy demand (P_{t}^{load}) using PV power (P_{t}^{pv}), buying power from the utility grid (P_{t}^{buy}), and/or by discharging the BESS (P_{t}^{disc}). Similarly, excess

power can be sold to the grid (P_t^{sell}). Finally, the BESS can be charged (P_t^{char}) either by buying power from the grid or using PV power.

$$P_t^{\text{buy}} + P_t^{\text{disc}} + P_t^{\text{pv}} = P_t^{\text{sell}} + P_t^{\text{char}} + P_t^{\text{load}}. \tag{5}$$

3. Soft Actor-Critic-Based Agent Training

In DRL, a deep neural network is combined with a reinforcement learning model, such as the Q model. This integrated model is capable of dealing with continuous state space and environments with uncertainty, which is not possible with Q-learning alone [42]. Several DRL variants have been described in the literature, but SAC is a state-of-the-art approach that has several advantages over other DRL methods. They include stable operation, fast convergence, and immunity to local optima trapping [43]. Therefore, SAC is used in this study to analyze the performance of PSM. SAC is used for both the conventional and the proposed methods in this study.

In SAC, the objective function is modified by adding an entropy term to measure the predictability of the random variable. Higher entropy values refer to lower predictability and vice versa. A hyperparameter α is used to control the trade-off between exploration and exploitation [44]. The agent learns a policy network (π_θ) and two Q networks (Q_1, Q_2). The two networks are trained independently, and two target Q-networks are also updated at each learning step via a soft copy update. The objective of the SAC agent is to maximize the sum of future expected rewards and future entropy (expected).

3.1. Demand Management of Buildings with Energy Storage and Solar PV

The demand management problem of buildings with BESS and PV is realized using the SAC method, due to its merits, as discussed in the previous paragraphs. The problem is modeled as a Markov decision process, and it includes state, action, state transition, and reward. The state at time t (S_t) contains information on the building load (P_t^{load}), PV power (P_t^{pv}), battery SoC (SoC_t), and interval number (t), as given below

$$S_t = [P_t^{\text{load}}, P_t^{\text{pv}}, SoC_t, t]. \tag{6}$$

The state parameters are normalized using minimum ($S^{\min}[k]$) and maximum ($S^{\max}[k]$) values of the states for fast convergence and stability of the learning process using

$$S_t[k] = \frac{S_t[k] - S^{\min}[k]}{S^{\max}[k] - S^{\min}[k]}, \tag{7}$$

where k is an index to access all four state parameters. Based on the current state of the system, the agent takes an action at time t (a_t) in the range of [-1,1] where negative values refer to BESS discharging and positive values refer to BESS charging. The actions are then transformed to the SoC level using the following equation

$$a_t^{\text{real}} = \frac{a_t}{M} \text{ where } M \in Q, \tag{8}$$

where the real actions (a_t^{real}) are obtained considering the charging/discharging rate of the BESS converter using the positive rational number M. For example, $M = 1$ refers to a C-rate of 1C and $M = 2$ refers to a C-rate of 0.5C, and so on.

The state transition from time t to $t+1$ depends on the values of building load, PV power, SoC level (state parameter at time t), and action taken at t. This can be mathematically modeled as

$$S_{t+1} = f(S_t, a_t). \tag{9}$$

Finally, the reward (r_t) is calculated using a reward function. Details about the reward function are discussed in the following sub-section.

3.1.1. Reward Function

A comprehensive reward function (r_t) is devised to manage the load of the building with BESS and PV. The objective of the agent is to flatten the load curve of the building without violating the BESS energy constraints [45]. The average load of each day (P^{avg}) is taken as a threshold, and the agent is rewarded for bringing the total load close to the threshold level. The reward function comprises three parts, as shown below

$$r_t = -\frac{(P^{\text{avg}} - (P_t^{\text{net}} + P_t^{\text{bess}}))^2}{P^{\text{avg}}} - p_1 - p_2. \tag{10}$$

The first part penalizes any deviations from the threshold level (daily average load). It is normalized with the daily average load value to scale the daily rewards to a similar level. The net load (P_t^{net}) in the first part refers to the load of the building after consuming PV energy, if available. The value of the net load is the same as the original load in the case of buildings without PV. It can be modeled as

$$P_t^{\text{net}} = \begin{cases} P_t^{\text{load}} - P_t^{\text{pv}}, & \text{if building contains PV,} \\ P_t^{\text{load}}, & \text{otherwise.} \end{cases} \tag{11}$$

Similarly, the variable P_t^{bess} refers to the power charged/discharged to/from the BESS at time t. It will take a positive value when BESS is charging and a negative value when BESS is discharging. It can be modeled as

$$P_t^{\text{bess}} = P_t^{\text{char}} - P_t^{\text{disc}}. \tag{12}$$

The last two terms of the reward function represent penalties for violating the SoC constraints. For example, if the BESS is overcharged (charged beyond the predefined upper limit), a penalty (p_1) of λ_1 is imposed as shown in the following equation

$$p_1 = \begin{cases} \lambda_1, & \text{if } SoC_t > SoC^{\max}, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Similarly, a penalty (p_2) of λ_2 is imposed if the BESS is subjected to deep-discharge (discharged below the lower limit). This can be modeled as

$$p_2 = \begin{cases} \lambda_2, & \text{if } SoC_t < SoC^{\min}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

It should be noted that all terms of the reward function have negative signs. It implies that all the terms are penalties and the objective of the DRL agent is to minimize these values (maximize the reward).

3.1.2. BESS Parameters

As mentioned in the previous section, the action of the agent (a_t) is transformed into BESS SoC. The SoC in the current interval t is updated based on the SoC of the previous interval (SoC_{t-1}) and the action taken by the agent, i.e., ΔSoC_t. Mathematically, this can be represented as

$$SoC_t = SoC_{t-1} + \Delta SoC_t, \Delta SoC_t \in [-1, 1]. \tag{15}$$

Then, the amount of power charged/discharged to/from the battery (P_t^{bess}) can be calculated based on the SoC at time t and $t-1$. This can be mathematically written as

$$P_t^{\text{bess}} = \frac{B^{\text{cap}} \times (SoC_t - SoC_{t-1})}{100}, \tag{16}$$

where B^{cap} is the capacity of the BESS in kWh. The amount of power charged to the BESS during the time interval t (P_t^{char}) can be computed using the following equation

$$P_t^{\text{char}} = \begin{cases} P_t^{\text{bess}}, & \text{if } P_t^{\text{bess}} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Finally, the following equation is used to determine the amount of power discharged from the battery in the time interval t (P_t^{disc})

$$P_t^{\text{disc}} = \begin{cases} -P_t^{\text{bess}}, & \text{if } P_t^{\text{bess}} < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

3.2. Diversity in Experience-Based Selection and Training

An overview of the PSM for DRL agents is presented in Section 2. In this section, the details of the training process are presented. The step-by-step process for training the DRL agent is shown in Algorithm 2. The inputs of the proposed algorithm are the (cluster-wise) arranged daily load profiles of the building. First, the optimal number of clusters (K) is obtained from Algorithm 1. Then, the total number of days in the data set (D) is calculated and the maximum number of episodes (N) for training the agent is set.

The first loop is used to determine the number of days in each cluster $i \in K$. Then the day ratio (R_i) of each cluster is determined based on the number of days in that cluster (D_i) and the total number of days in the data set ($D = \sum_{i=1}^{K} D_i$). This can be mathematically represented as

$$R_i = \frac{D_i}{\sum_{i=1}^{K} D_i} \times 100. \tag{19}$$

Algorithm 2: PSM training process.

Data: Arranged daily energy consumption profiles of the building: $[D_1, D_2, ..D_i]$
Get the optimal number of clusters (K) from **Algorithm 1**
Get the total number of days in the training set (D)
Set the total number of episodes for training (N)
for *i=1:K* **do**
    Determine the number of days in cluster i (D_i)
    Determine the day ratio for cluster i (R_i) using (19)
    Compute the cumulative ratio for each cluster i (R_i^c) using (20)
end
for *j=1:N* **do**
    if MOD$(j, 100) > 0$ AND $\leq R_1^c$ **then**
        Select a random day in the range [1,D_1]
    else if MOD *(j,100)* $> R_1^c$ AND $\leq R_2^c$ **then**
        Select a random day in the range [D_1,D_2]
    else if MOD *(j,100)* $> R_{i-2}^c$ AND $\leq R_{i-1}^c$ **then**
        Select a random day in the range [D_{i-2},D_{i-1}]
    else
        Select a random day in the range [D_{i-1},D_i]
    end
end
Result: Daily energy consumption profiles based on equal experience

Then, the cumulative day ratio (R_i^c) for each cluster is determined by summing the day ratio of all clusters from 1 to the current cluster i. This can be mathematically written as

$$R_i^c = (\sum_{k=1}^{i} R_k). \tag{20}$$

The second loop is then used to iterate over the number of episodes (N). For each episode j, a sample needs to be selected from one of the clusters. The selection of the sample from any cluster is based on the value of the cumulative ratio computed using (20). For example, the first R_1^c samples will be selected from cluster 1, R_2^c samples from cluster 2, and so on. This process is repeated every 100 episodes and continues until the end of the episodes.

3.3. Indices for Performance Analysis

To analyze the performance of the selection method, two indices are used in this study. The same indices are used to compare the performance of the PSM with the RSM. The first index is called flatness (FI) and measures the flatness of the energy demand profile. It can be mathematically modeled using the daily maximum load ($P^{\max}$) and the daily minimum load ($P^{\min}$)

$$\text{FI} = 100 - (\frac{P^{\max} - P^{\min}}{P^{\max}} \times 100). \tag{21}$$

It is a continuous index that ranges between [0,100] for buildings without PV. Higher values correspond to higher flatness and vice versa. It should be noted that the value of FI could be negative for buildings with PV.

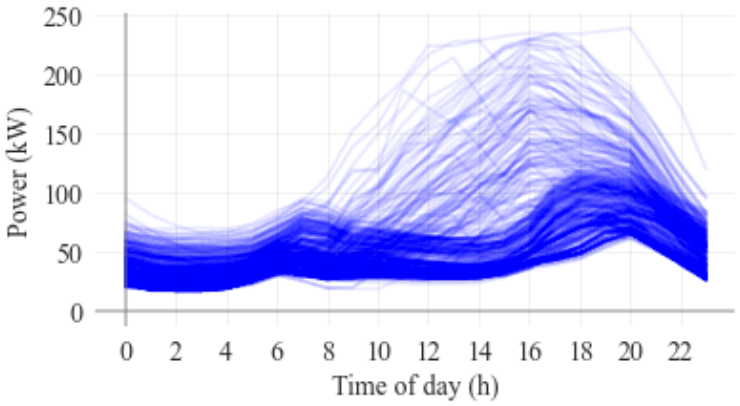

Figure 3: Overview of raw data used in this study.

The second index is called divergence (DI) and measures the divergence of the net load (P_t^{net}) from the daily average load (P^{avg}). It can be mathematically written as

$$\text{DI} = \frac{\sqrt{\sum_{t=1}^{T}(P_t^{\text{net}} - P^{\text{avg}})^2}}{T} \tag{22}$$

Similar to the previous index, it is also a continuous measure, where higher values refer to higher divergence and vice versa. Note that higher values of FI and lower values of DI are desired.

4. Performance Evaluation

In this section, the raw data is preprocessed and grouped into different clusters using the previously discussed methods. Then, the performance of the PSM is compared with that of the RSM in terms of training the DRL agent. Both the proportional and conventional DRL methods are implemented in Python 3.9.13.

4.1. Input Data

The input data comprises the real energy consumption of a residential complex in the USA [46]. The data used in this study have a resolution of 1 hour over 6 consecutive years. It is a scaled version of the data reported in [46]. PV power data for the corresponding years are generated using [47]. An overview of the raw data, showing different load profile groups, is shown in Fig. 3. For the sake of visualization, selected PV power and load profiles of different days and seasons are shown in Fig. 4. PV power shows higher energy (longer days and higher peaks) during summers and lower energy (shorter days and lower peaks) during winters, as expected. Similarly, the building load profiles are significantly different for different seasons of the year. In addition, the differences between weekdays and holidays are also evident for the same seasons. This implies that the BESS operation needs to be adjusted for different seasons of the year and days of the week.

The BESS of 250kWh is selected considering the load and PV profiles of the building. The upper and lower bounds of the SoC are taken as 90% and 10%, respectively. Finally, the

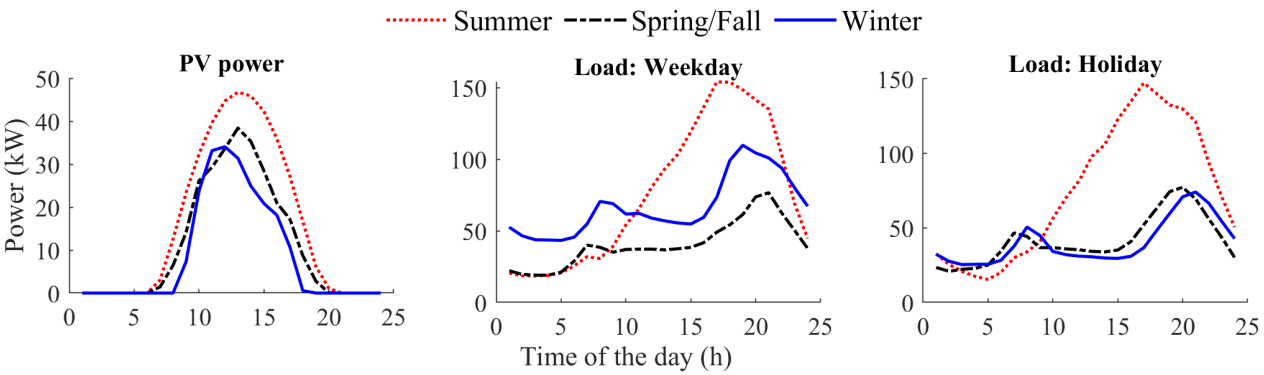

Figure 4: Daily load profiles of selected days for different day types and seasons.

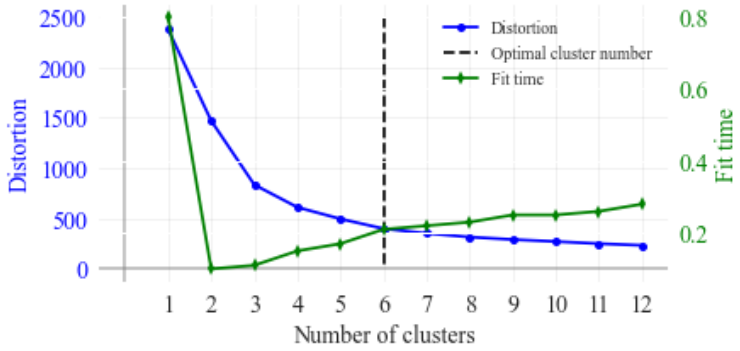

Figure 5: Determining the optimal number of clusters using the Elbow method.

charging and discharging rates of the BESS are taken as C/5, which means that the BESS can be fully charged or discharged in 5 hours.

4.2. Analysis of Optimal Number of Clusters

In this section, Algorithm 1 is used to determine the optimal number of clusters to group the daily load profiles into different groups. The number of clusters is varied between 1 to 12 and distortion and computation time (fit time) are recorded for each run. The recorded data are plotted in Fig. 5 and it can be observed that the distortion decreases with an increase in the number of clusters. Because distortion measures the difference between the data sample and its centroid, it is expected that with more centroids (clusters), the distortion will decrease. However, the rate of decrease in distortion is not uniform; i.e., it initially decreases rapidly and then it flattens out. The elbow method is used to identify the point where this transition occurs to determine the optimal number of clusters [48]; in this case $K = 6$. Similarly to [49], the fit time is also shown in Fig. 5. It shows an initial decrease, and then it increases with an increase in the number of clusters.

After identifying the optimal number of clusters, the raw data (shown in Fig. 3) are grouped into different clusters. Each cluster is represented with a different color, as shown in Fig. 6. The dashed lines in each color represent the centroids of each cluster.

Generally, dimensionality reduction methods are used to validate the clustering results. The t-distributed stochastic neighbor embedding (t-SNE) [50] is a popular and widely used

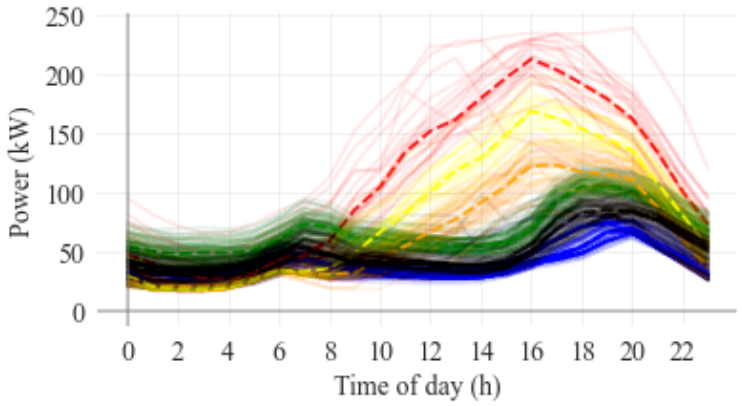

Figure 6: Grouping of load profiles into different clusters.

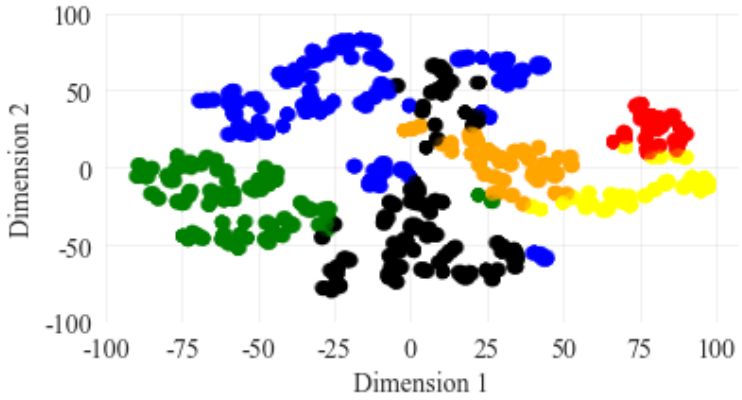

Figure 7: Overview of results obtained from the t-SNE method with two dimensions.

method [51]. Therefore, this method is also used in this study to reduce the dimension from 24 to 2. The obtained results are shown in Fig. 7, where each dot represents a daily load profile. The colors correspond to the clusters shown in Fig. 6. The presence of similar color dots close to one another shows that the K-means method has successfully grouped the data into different clusters.

The summary of data points in each cluster (D_i), the day ratio of each cluster (R_i), and the cumulative day ratio of each cluster (R_i^c) are shown in Table 1. Furthermore, the prevailing time period of the year (including the season and type of day) for each cluster is also indicated in the table.

Table 1: Summary of data samples and related statistics in each cluster.

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Time Period	Summer	Weekdays Spring/Fall	Winter	Summer	Holidays Spring/Fall	Winter
D_i	510	370	625	214	164	309
$R_i(\%)$	23	17	29	10	8	14
$R_i^c(\%)$	23	40	69	78	86	100

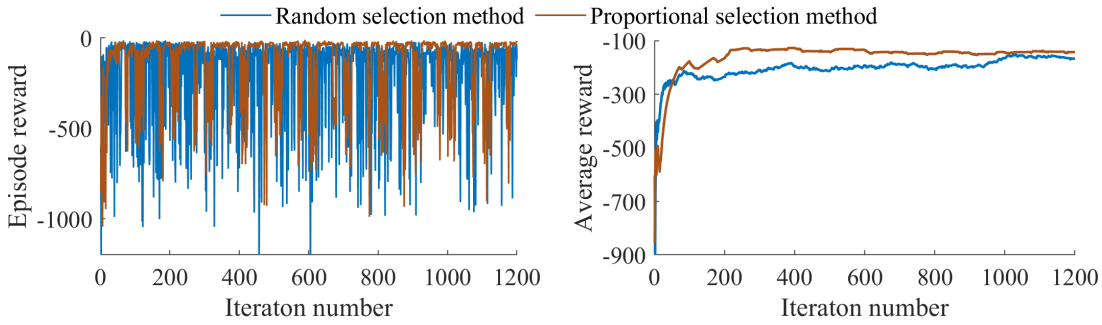

Figure 8: Convergence and stability results with 1-year data.

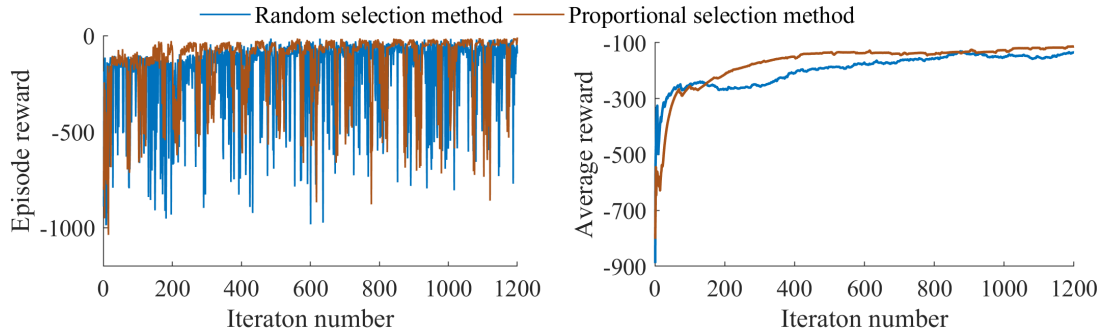

Figure 9: Convergence and stability results with 3-years data.

4.3. Convergence and Stability Analysis

The DRL agent is trained using data of 5 years (as discussed in the previous section). Data from the 6^{th} year are used to validate performance, as explained in the next section. To analyze the performance of the proposed method, the model is trained with different datasets, i.e., 1 year, 3 years, and 5 years without including PV (only BESS). Finally, an additional case is simulated that includes both BESS and PV in the building.

The convergence and stability analysis results of the first three cases (1 year, 3 years, and 5 years data) are shown in Fig. 8, Fig. 9, and Fig. 10, respectively. The left sides of the figures show the per-episode rewards, and the right sides represent the running average rewards of 100 episodes. It can be seen that the PSM has outperformed the RSM in all three cases. The performance of the PSM is superior to that of the RSM in terms of average reward, stability, and convergence speed. The PSM has converged to higher rewards in all cases, which implies that it has well-explored the environment and found better actions (BESS charging/discharging rates and intervals). In addition, its average reward has lower fluctuations compared to the conventional method, indicating better stability of the learning process. Finally, the PSM converged faster in all cases. Especially, the performance of the RSM has shown a deteriorating trend with the increased number of data points, i.e., the 5-years case.

The results for the building with BESS and PV are shown in Fig. 11, where the agent is trained using the 3-year data set. A similar trend can be observed in this case, similar to the previous three. The PSM has outperformed the RSM in terms of stability, convergence speed, and reward levels. This analysis shows that the proposed approach performs better

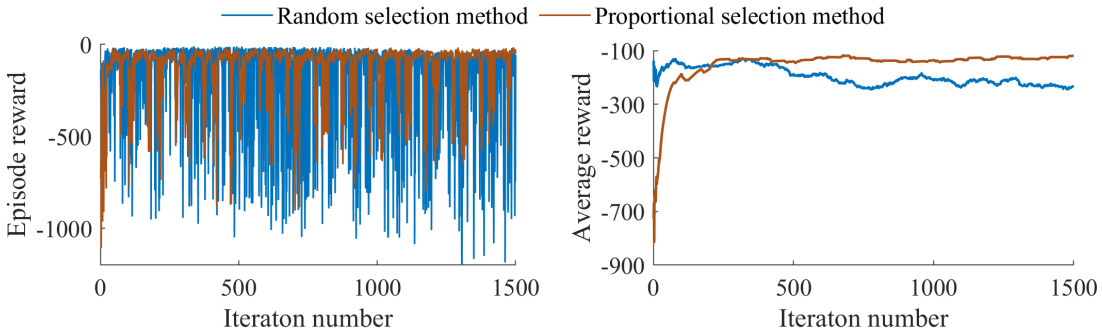

Figure 10: Convergence and stability results with 5-years data.

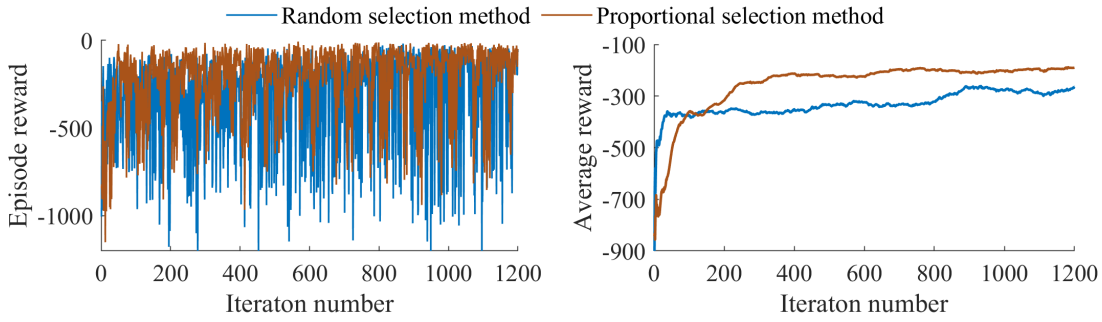

Figure 11: Convergence and stability results with inclusion of PV.

in different environments as compared to the RSM. The actions that resulted in different rewards are compared and contrasted for both methods in the following sections.

5. Discussion and Analysis

In this section, the performance of the PSM is compared with the RSM using the two indices formulated in Section 3.3. Different scenarios are considered to analyze the performance of the PSM such as different data sets, integration of PV, and daily profile analysis.

5.1. Analysis with Different Datasets

In this section, three data sets are considered to analyze the performance of the PSM and the conventional RSM for the data shown in Fig. 4. The results of the flatness and divergence indices with 1-year data are shown in Fig. 12. Similarly, Figs. 13 and 14 show the results with 3-year and 5-year data, respectively.

It can be observed that the flatness index is the lowest for the original case in all scenarios, as expected. The original case refers to the original load without BESS integration. The flatness index has also increased for the RSM case, as compared to the original case, due to the integration of the BESS. However, in all scenarios, the flatness index is the highest for the PSM. Higher values of the flatness index refer to more even load curves, which is desired. The highest difference in the flatness index was about 32% compared to the RSM (cluster 2) and about 63% compared to the original method (cluster 4).

The divergence index analysis shows that the PSM has the lowest values in all scenarios. Higher values of DI refer to higher divergence, and thus lower values are desirable. The highest

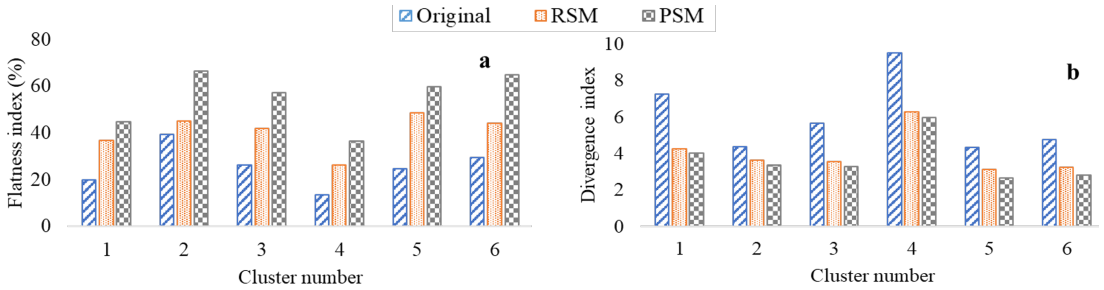

Figure 12: Load variation results with 1-year data: a) flatness index; b) divergence index.

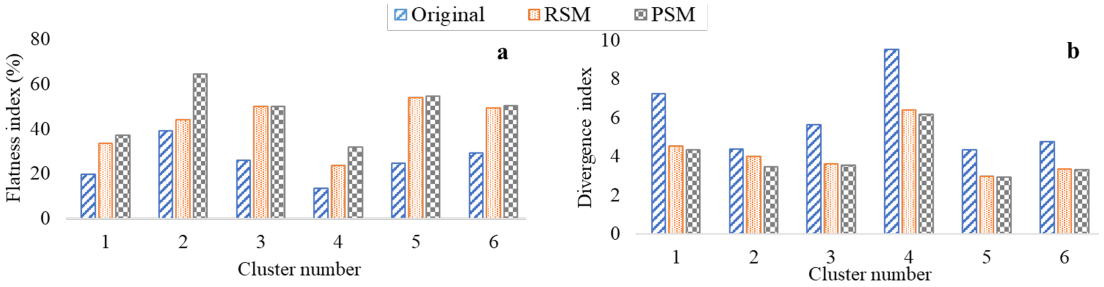

Figure 13: Load variation results with 3-year data: a) flatness index; b) divergence index.

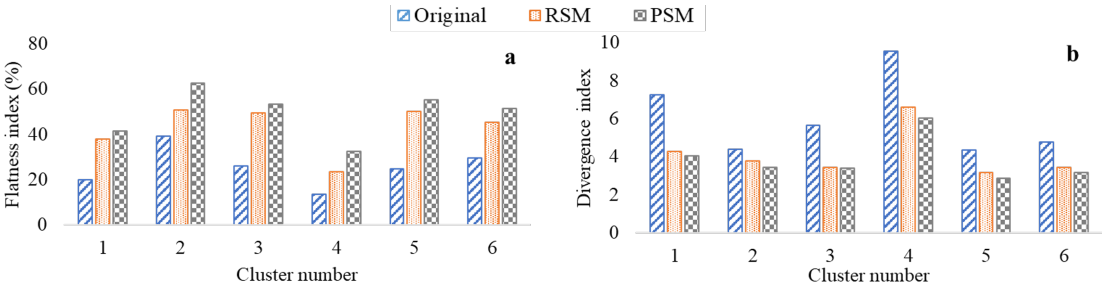

Figure 14: Load variation results with 5-year data: a) flatness index; b) divergence index.

difference was about -15% compared to the RSM (cluster 5) and about -45% compared to the original case (cluster 1).

This analysis shows that the PSM has superior performance compared to the conventional RSM in terms of both flatness and divergence. This is in accordance with the convergence analysis, where the PSM has always converged to a higher reward corresponding to better utilization of BESS in flattening the load curve.

5.2. Analysis with Integration of PV

In this section, the performance of the PSM is compared to that of the conventional RSM, while considering the integration of PV. Similar to the previous section, the flatness and divergences indices are analyzed. It can be observed from Table 2 that the flatness index is the highest for all cases for the PSM, which is desired. Similarly, divergence analyses in Table 3 show that the divergence index is the lowest in all cases for the proposed approach. A lower value of DI is desired, since it refers to a lower divergence from the daily average load. This analysis shows that the proposed approach has superior performance even with

the integration of PV in the building.

Table 2: Flatness index results (%) with the integration of BESS and PV.

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Original	19.7	32.5	39.9	15.7	40.2	41.6
With PV	-6.3	14.2	19.0	13.9	9.4	11.3
RSM	17.4	32.4	44.4	12.4	45.8	43.7
PSM	29.9	64.2	57.9	20.9	55.0	50.7

Table 3: Divergence index results with the integration of BESS and PV.

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Original	7.3	4.1	4.5	8.2	4.0	5.2
With PV	8.5	4.9	5.3	8.8	5.4	6.7
RSM	5.5	3.5	3.9	7.5	3.8	4.4
PSM	5.1	2.9	3.4	6.3	3.5	4.2

5.3. Profile Comparisons

In this section, the daily load profiles are analyzed for selected days of the year (one from each cluster). For the sake of visualization, only one case (3 years of data) from different datasets is presented, Figs. 15 and 16. In addition, the results of the PV-integrated case are also analyzed, Figs. 17 and 18. The results presented in this section are obtained using the trained SAC agent. As discussed in Section 3.1, each state encompasses information on the building load (P_t^{load}), PV power (P_t^{pv}), battery SoC (SoC_t), and the interval number (t). Likewise, the action taken by the agent (a_t) is converted into battery SoC using equation (8). This SoC value is then used to determine the quantum of power charged/discharged to/from the BESS during the interval t using equation (16). Additionally, the agent is equipped with penalties for exceeding SoC limits, as elaborated in Section 3.1.1.

To achieve the outcomes discussed in this section, the trained agent is provided with load and PV power data for a 24-hour period (one day) for each cluster individually. The agent's interval-specific SoC (action) is recorded and utilized to calculate the battery power. This methodology leads to the results depicted in Figures 16 and 18. Similarly, adjusted load profiles are obtained by combining the original building load (P_t^{load}), PV power (P_t^{pv}), and battery power SoC (P_t^{bess}), i.e., P_t^{load} - P_t^{pv} + P_t^{bess}. The adjusted load profiles of PSM and RSM, shown in Figures 15 and 17, are obtained using this approach. It should be noted that P_t^{bess} will exhibit a positive value during charging and a negative value during discharging. Consequently, the load will increase during charging intervals and decrease during discharging intervals.

The daily load profiles for the original case, conventional RSM and PSM are shown in Fig. 15. The charging/discharging power and SoC of BESS are shown in Fig. 16. It should be noted that the positive values in Fig. 16 refer to the charging power and the negative values refer to the discharging power. The flatness of the load profiles in comparison to the original load profiles is obvious. The difference in the charging and discharging profiles for RSM and PSM can be observed in Fig. 16, which resulted in the difference in the load profiles,

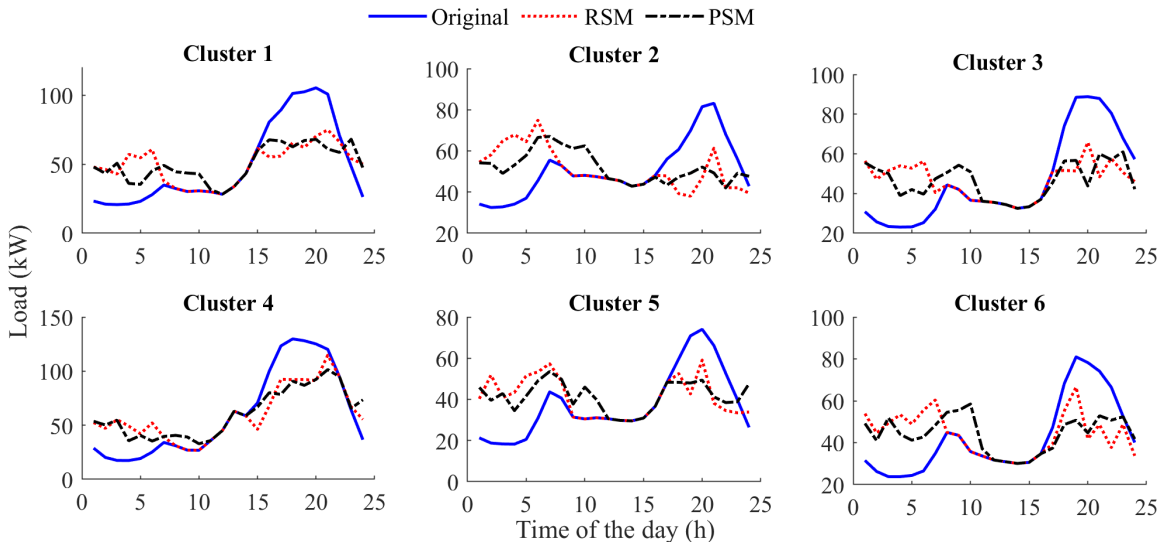

Figure 15: Load profiles of selected days with model training using 3 years of data.

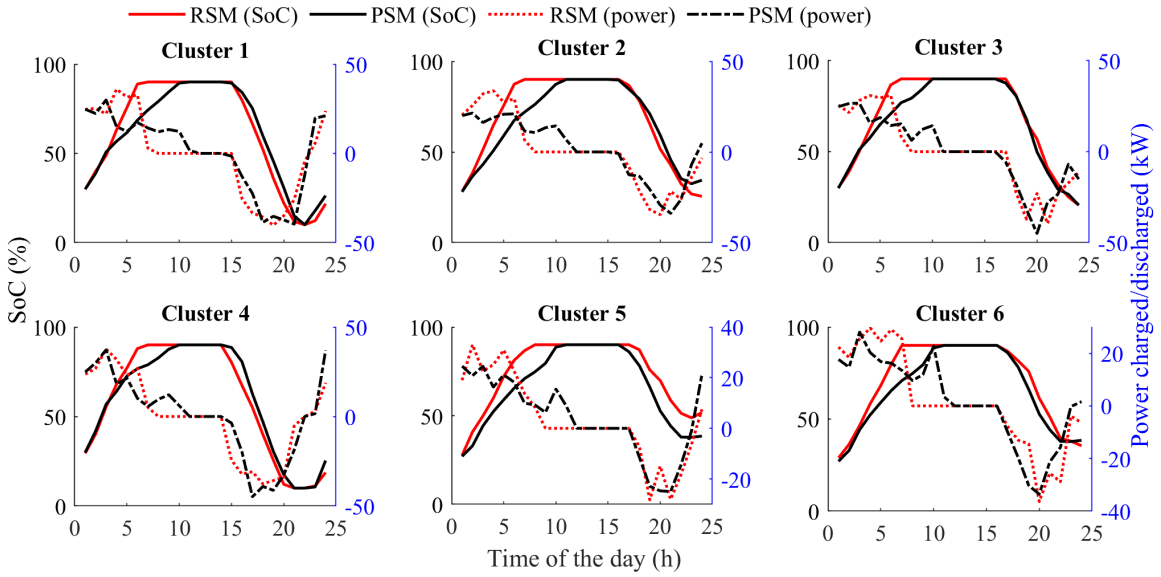

Figure 16: BESS power and SoC profiles with model training using 3 years of data.

shown in Fig. 15. The proposed approach has reached the maximum SoC level later in all cases, as compared to the RSM. This means that BESS has charged over a larger number of intervals (using lower power in each interval), resulting in flatter load curves. It can also be observed that the difference in peak-to-valley load is lower for PSM, compared to RSM, in all cases. The term "peak" denotes the maximum value within the load profile, while "valley" refers to the minimum value. Consequently, "peak-to-valley" signifies the disparity between the highest and lowest load points. Reduced peak-to-valley values are preferred as they indicate a more even load distribution. The outcomes of various indices (such as the flatness index and divergence index) have substantiated these findings, as evidenced in the preceding section, 5.1.

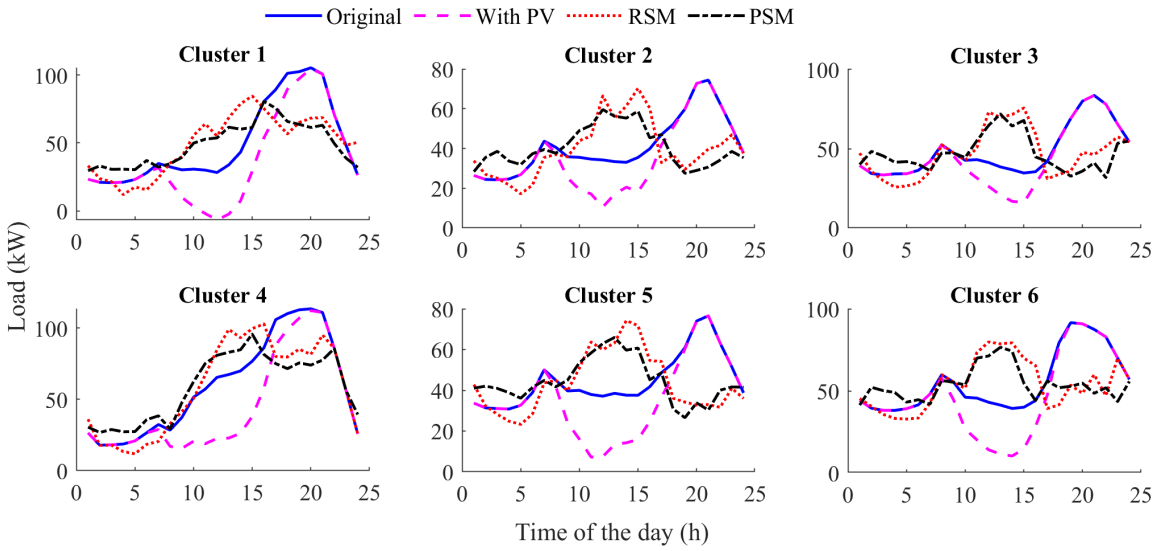

Figure 17: Load profiles of selected days with the integration of PV.

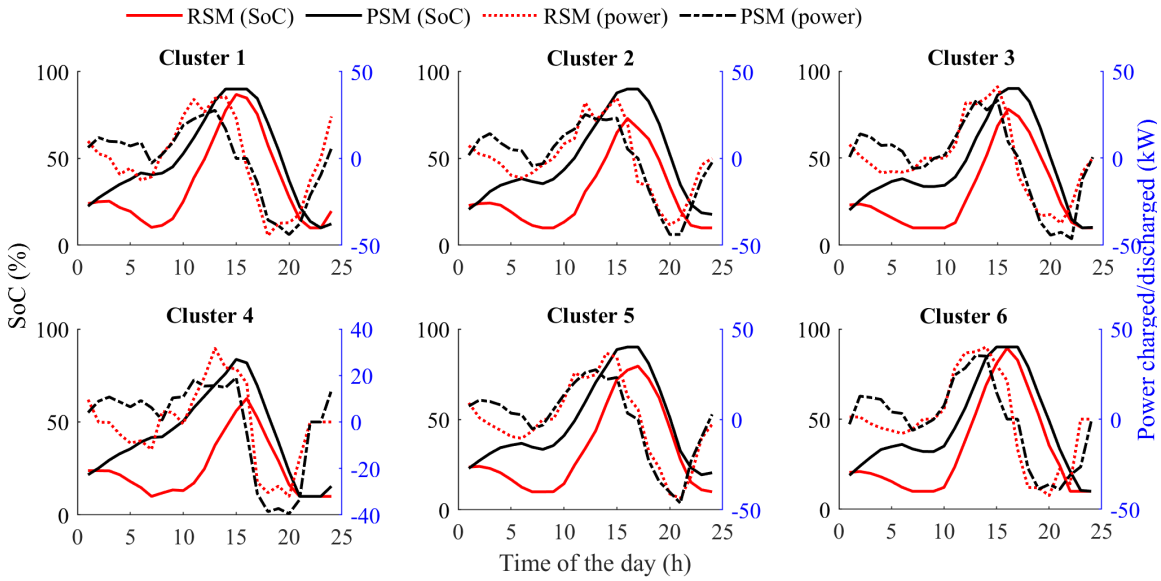

Figure 18: BESS power and SoC profiles with the integration of PV.

The load profiles for the PV-integrated case are shown in Fig. 17, and the corresponding BESS charging/discharging profiles and SoC are shown in Fig. 18. It can be observed that the load profiles are much flatter with the integration of the BESS as compared to the original case and the case with PV. It should be noted that the net load (with PV) has become negative in Cluster 1. In addition, the load with a PV profile has a higher peak-to-valley difference compared to the original load. The proposed approach has resulted in flatter load curves as compared to the RSM. It can also be observed that the RSM was not able to fully utilize the BESS in most cases (not charged to the upper SoC bound). In contrast, the proposed approach has fully utilized BESS in all cases. This also resulted in the inferior performance of the RSM as evaluated using different indices in the previous section.

6. Conclusions

A deep reinforcement learning model based on diversity in experience is proposed for training agents to manage the load of buildings with energy storage and solar PV. The raw data are grouped into different clusters and a selection algorithm is devised to select data from each cluster to ensure the exposure of the agent to all types of data. Convergence analysis has shown that the proposed approach has faster convergence compared to the approach using the conventional random selection method. In addition, the proposed model has always converged to a higher reward and with better stability for average reward. It implies that the proposed approach can better explore the environment and find optimal actions to manage the load by controlling the battery charging/discharging. Similarly, the flatness index of the proposed approach was higher for all cases and the difference was up to 32% lower compared to the conventional approach. The divergence index was the lowest for the proposed approach in all cases, indicating that the resulting load profile is close to the daily average load. The difference in the divergence index was up to 15% compared to the conventional method. A detailed analysis has shown that the proposed approach has superior performance as a result of its ability to explore representative data profiles from the entire environment and choose better actions for battery charging/discharging.

References

[1] J. Liu, X. Chen, H. Yang, Y. Li, Energy storage and management system design optimization for a photovoltaic integrated low-energy building, Energy 190 (2020) 116424.

[2] Global status report for buildings and construction, Tech. rep., United Nations Environment Programme, https://www.unep.org/resources/publication/2022-global-status-report-buildings-and-construction (2022).

[3] Electricity market report 2023, Tech. rep., IEA, https://www.iea.org/events/electricity-market-report-2023 (2023).

[4] Aeso net-zero emissions pathways, Tech. rep., AESO, Canada, https://www.aeso.ca/market/net-zero-emissions-pathways (2022).

[5] A. Hussain, P. Musilek, Resilience enhancement strategies for and through electric vehicles, Sustainable Cities and Society (2022) 103788.

[6] M. Wei, C. A. McMillan, S. de la Rue du Can, Electrification of industry: Potential, challenges and outlook, Current Sustainable/Renewable Energy Reports 6 (2019) 140–148.

[7] P. R. White, J. D. Rhodes, E. J. Wilson, M. E. Webber, Quantifying the impact of residential space heating electrification on the texas electric grid, Applied Energy 298 (2021) 117113.

[8] C.-Y. Chen, K. K. Chai, E. Lau, Ai-assisted approach for building energy and carbon footprint modeling, Energy and AI 5 (2021) 100091.

[9] M. Stratistics, Uav drones–global market outlook (2016–2022), Report ID: SMRC16075 (2016).

[10] S. Bahramara, Robust optimization of the flexibility-constrained energy management problem for a smart home with rooftop photovoltaic and an energy storage, Journal of Energy Storage 36 (2021) 102358.

[11] J. Niu, Z. Tian, Y. Lu, H. Zhao, Flexible dispatch of a building energy system using building thermal storage and battery energy storage, Applied Energy 243 (2019) 274–287.

[12] J. Liu, X. Chen, S. Cao, H. Yang, Overview on hybrid solar photovoltaic-electrical energy storage technologies for power supply to buildings, Energy conversion and management 187 (2019) 103–121.

[13] T. Terlouw, T. AlSkaif, C. Bauer, W. van Sark, Optimal energy management in all-electric residential energy systems with heat and electricity storage, Applied Energy 254 (2019) 113580.

[14] J. Koskela, A. Rautiainen, P. Järventausta, Using electrical energy storage in residential buildings–sizing of battery and photovoltaic panels based on electricity cost optimization, Applied energy 239 (2019) 1175–1189.

[15] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, F. S. García, A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis, Journal of Building Engineering 33 (2021) 101692.

[16] P. W. Tien, S. Wei, J. Darkwa, C. Wood, J. K. Calautit, Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality–a review, Energy and AI (2022) 100198.

[17] H. Golpîra, S. A. R. Khan, A multi-objective risk-based robust optimization approach to energy management in smart residential buildings under combined demand and supply uncertainty, Energy 170 (2019) 1113–1129.

[18] M. S. Javadi, M. Gough, M. Lotfi, A. E. Nezhad, S. F. Santos, J. P. Catalão, Optimal self-scheduling of home energy management system in the presence of photovoltaic power generation and batteries, Energy 210 (2020) 118568.

[19] C. L. Nge, I. U. Ranaweera, O.-M. Midtgård, L. Norum, A real-time energy management system for smart grid integrated photovoltaic generation with battery storage, Renewable energy 130 (2019) 774–785.

[20] B. Lokeshgupta, S. Sivasubramani, Multi-objective home energy management with battery energy storage systems, Sustainable Cities and Society 47 (2019) 101458.

[21] S. Ahmad, M. Naeem, A. Ahmad, Low complexity approach for energy management in residential buildings, International Transactions on Electrical Energy Systems 29 (1) (2019) e2680.

[22] D. Azuatalam, W.-L. Lee, F. de Nijs, A. Liebman, Reinforcement learning for whole-building hvac control and demand response, Energy and AI 2 (2020) 100020.

[23] S. Abedi, S. W. Yoon, S. Kwon, Battery energy storage control using a reinforcement learning approach with cyclic time-dependent markov process, International Journal of Electrical Power & Energy Systems 134 (2022) 107368.

[24] F. Alfaverh, M. Denai, Y. Sun, Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management, IEEE access 8 (2020) 39310–39321.

[25] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, Deep reinforcement learning for smart home energy management, IEEE Internet of Things Journal 7 (4) (2019) 2751–2762.

[26] R. Shen, S. Zhong, X. Wen, Q. An, R. Zheng, Y. Li, J. Zhao, Multi-agent deep reinforcement learning optimization framework for building energy system with renewable energy, Applied Energy 312 (2022) 118724.

[27] G. Pinto, D. Deltetto, A. Capozzoli, Data-driven district energy management with surrogate models and deep reinforcement learning, Applied Energy 304 (2021) 117642.

[28] G. Pinto, M. S. Piscitelli, J. R. Vázquez-Canteli, Z. Nagy, A. Capozzoli, Coordinated energy management for a cluster of buildings through deep reinforcement learning, Energy 229 (2021) 120725.

[29] A. Shaqour, A. Hagishima, Systematic review on deep reinforcement learning-based energy management for different building types, Energies 15 (22) (2022) 8663.

[30] Z. Wang, T. Hong, Reinforcement learning for building controls: The opportunities and challenges, Applied Energy 269 (2020) 115036.

[31] Z. Nagy, G. Henze, S. Dey, J. Arroyo, L. Helsen, X. Zhang, B. Chen, K. Amasyali, K. Kurte, A. Zamzam, et al., Ten questions concerning reinforcement learning for building energy management, Building and Environment (2023) 110435.

[32] D. Weinberg, Q. Wang, T. O. Timoudas, C. Fischione, A review of reinforcement learning for controlling building energy systems from a computer science perspective, Sustainable cities and society (2022) 104351.

[33] K. Nweye, B. Liu, P. Stone, Z. Nagy, Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings, Energy and AI 10 (2022) 100202.

[34] J. Parker-Holder, A. Pacchiano, K. M. Choromanski, S. J. Roberts, Effective diversity in population based reinforcement learning, Advances in Neural Information Processing Systems 33 (2020) 18050–18062.

[35] S. Dey, T. Marzullo, X. Zhang, G. Henze, Reinforcement learning building control approach harnessing imitation learning, Energy and AI 14 (2023) 100255.

[36] Z.-W. Hong, T.-Y. Shann, S.-Y. Su, Y.-H. Chang, T.-J. Fu, C.-Y. Lee, Diversity-driven exploration strategy for deep reinforcement learning, Advances in neural information processing systems 31 (2018).

[37] T. Pereira, M. Abbasi, B. Ribeiro, J. P. Arrais, Diversity oriented deep reinforcement learning for targeted molecule generation, Journal of cheminformatics 13 (2021) 1–17.

[38] T. Dai, Y. Du, M. Fang, A. A. Bharath, Diversity-augmented intrinsic motivation for deep reinforcement learning, Neurocomputing 468 (2022) 396–406.

[39] K. Kanishev, Pyidf: Diversity of experiences in reinforcement learning, https://medium.com/imandra/pyidf-diversity-of-experiences-in-reinforcement-learning-8d59f60f59ed (2019).

[40] M. Nazari, A. Hussain, P. Musilek, Applications of clustering methods for different aspects of electric vehicles, Electronics 12 (4) (2023) 790.

[41] P. Govender, V. Sivakumar, Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019), Atmospheric pollution research 11 (1) (2020) 40–56.

[42] A. Hussain, V.-H. Bui, H.-M. Kim, Deep reinforcement learning-based operation of fast charging stations coupled with energy storage system, Electric Power Systems Research 210 (2022) 108087.

[43] V.-H. Bui, W. Su, Real-time operation of distribution network: A deep reinforcement learning-based reconfiguration approach, Sustainable energy technologies and assessments 50 (2022) 101841.

[44] V.-H. Bui, A. Hussain, W. Su, A dynamic internal trading price strategy for networked microgrids: A deep reinforcement learning-based game-theoretic approach, IEEE Transactions on Smart Grid 13 (5) (2022) 3408–3421.

[45] A. Hussain, P. Musilek, Utility-scale energy storage system for load management under high penetration of electric vehicles: A marginal capacity value-based sizing approach, Journal of Energy Storage 56 (2022) 105922.

[46] S. Taheri, M. Jooshaki, M. Moeini-Aghtaie, 8 years of hourly heat and electricity demand for a residential building (2021). doi:10.21227/dfvb-re49.

[47] Renewables ninja, https://www.renewables.ninja.

[48] Elbow method for optimal value of k in kmeans, https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/.

[49] T. T. Teoh, Z. Rong, Clustering, Springer Singapore, Singapore, 2022, pp. 213–218.

[50] Ml — t-distributed stochastic neighbor embedding (t-sne) algorithm, https://www.geeksforgeeks.org/ml-t-distributed-stochastic-neighbor-embedding-t-sne-algorithm/.

[51] L. G. Viola, Clustering electricity usage profiles with k-means, https://towardsdatascience.com/clustering-electricity-profiles-with-k-means-42d6d0644d00.