Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Deep reinforcement learning-based operation of fast charging stations coupled with energy storage system

Akhtar Hussain [a], Van-Hai Bui [b], Hak-Man Kim [c,*]

[a] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2G2, Canada
[b] College of Engineering and Computer Science, University of Michigan-Dearborn, Dearborn, MI 48128, United States
[c] Department of Electrical Engineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, South Korea

ARTICLE INFO

Keywords:
Charging station
Deep reinforcement learning
Electric vehicles
Energy storage system
Reinforcement learning
Smart grid

ABSTRACT

Fast charging stations (FCSs) can reduce the charging time of electric vehicles (EVs) and thus can help in the widespread adoption of EVs. However, FCSs may result in the power system overload. Therefore, the deployment of the battery energy storage system (BESS) in FCSs is considered as a potential solution to avoid system overload. However, the optimal operation of FCSs equipped with BESS is challenging due to the involvement of several uncertainties, such as EV arrival/departure times and electricity prices. Therefore, in this study, a deep reinforcement learning-based method is proposed to operate FCSs with BESS under these uncertainties. The state-of-the-art soft actor-critic method (SAC) is adopted and the model is trained with one-year data to cover seasonality and different types of days (working days and holidays). The performance of SAC is compared with two other deep reinforcement learning methods, i.e., deep deterministic policy gradient and twin delayed deep deterministic policy gradient. A comprehensive reward function is devised to train the model offline, which can then be used for the real-time operation of FCS with BESS under different uncertainties. The trained model has successfully reduced the peak load of the FCS during both weekdays and holidays by optimizing the operation of the BESS. In addition, the robustness of the proposed model against different EV arrival scenarios and extreme market price scenarios is also evaluated. Simulation results have shown that the proposed model can reduce the peak load of the FCS under diverse conditions in the desired fashion.

Nomenclature

BESS	battery energy storage system
KOSIS	Korea statistical information service
CSO	charging station operator
PDF	probability density functions
DDPG	deep deterministic policy gradient
SAC	soft actor-critic
DRL	deep reinforcement learning
SOC	state-of-charge
EV	electric vehicle
TD3	twin delayed DDPG
FCS	fast charging station

1. Introduction

Penetration of electric vehicles (EVs) is increasing but several issues need to be resolved for widespread adoption. The three major barriers to the widespread adoption of EVs are the upfront cost of EVs, range anxiety, and absence of charging infrastructure [1]. The large size of EV batteries can solve the range anxiety, but it will result in the increased cost of EVs since the battery accounts for almost half of the EV cost [2]. Therefore, widespread deployment of charging stations is the second alternative to overcome the range anxiety issues. Especially, fast charging stations (FCS) are required to reduce the recharge time. Deployment of FCS can not only reduce the range anxiety but can also reduce recharging time and enhance the driving range. However, they pose several challenges to the power network, especially the power distribution system [3]. For example, overloading of local power equipment such as feeders, transformers, and distribution lines. The increased penetration of FCS may necessitate the reinforcement of grid infrastructure, which requires huge investments and time. Several smart charging methods [4–6] are proposed in the literature to cover this issue but these methods are only beneficial for those charging stations, where

[*] Corresponding author.
E-mail address: hmkim@inu.ac.kr (H.-M. Kim).

https://doi.org/10.1016/j.epsr.2022.108087
Received 14 November 2021; Received in revised form 16 February 2022; Accepted 5 April 2022
Available online 19 May 2022

there is enough time for shifting loads. In addition, only smart charging cannot solve the issue for clustered EVs, where several EVs are connected to a single charging port. Therefore, the use of a stationary battery energy storage system (BESS) is proposed as a potential alternative in the literature [7].

BESS can not only reduce the system peak load by charging EVs during peak intervals but can also be used for providing other services to the grid during idling time [8,9]. These additional services will increase the profit of charging station operators by generating additional revenues. Detailed analysis of potential benefits of FCSs with BESS can be found in [7–10], where techno-economic benefits are analyzed. To achieve the benefits of BESS, optimal sizing of BESS is required. Therefore, several studies are conducted on optimal sizing of BESS considering different objectives, which are as follows. In [11], sizing of BESS is carried out to minimize both yearly operational costs and the upfront investment cost. In addition to the system cost, the resilience of EVs during outages is considered in [12] and the size of stationary BESS is determined to feed the EVs during system contingencies. In addition to BESS, integration of renewables is also considered in several studies. For example, the capacity configuration of renewables and BESS [13], consideration of traffic uncertainty for determining sizes of BESS and renewables [14], and detailed modeling of renewables and degradation of BESS [15].

The BESS needs to operate optimally to reduce the daily peak load and to avoid system overloading. However, the operation of BESS in an FCS is subjected to several uncertainties, such as the arrival and departure times of EVs, the required energy by each EV, and the upstream grid prices. The required energy by each EV depends on the state-of-charge (SOC) at the beginning of the day and the daily mileage of the EV. Therefore, the required energy along with arrival and departure times of EVs are modeled as stochastic variables [12,13]. However, stochastic optimization requires strong assumptions on the uncertain parameters, and generally, probability density functions (PDFs) are used and it is difficult to determine accurate PDFs [16]. In addition, scenario generation and reduction methods are required and thus the performance is subjected to the accuracy of these underlying models as well [17]. Stochastic optimization suffers from the curse of dimensionality and computational complexity increases significantly with an increase in the problem/random variable size [18]. The stochastic optimization results in probabilistic results and it have been shown in several studies that learning methods perform better than the probabilistic methods [19]. Especially, reinforcement learning-based approaches are proven to be beneficial for systems with unknown dynamics or subjected to diverse uncertainties [20] due to the model-free nature of the method. Although only reinforcement learning may face issues with continuous action space, it can be covered by augmenting it with deep learning, known as deep reinforcement learning (DRL) [21]. Therefore, DRL is recently used for several problems related to power systems and electric vehicles.

The DRL studies on power systems can be categorized as studies on distribution systems, microgrids, charging stations, and electric vehicles. For example, DRL is applied for demand response in a distribution network considering uncertainty in electricity prices, users' comfort, and load. DRL is used to realize reactive power control in active distribution networks [22] and real-time reconfiguration of the distribution network [23]. Similarly, DRL is applied for different applications in microgrids. For example, the operation of an energy storage system under volatile energy sources is determined using DRL in [24]. Similarly, dynamic energy dispatch of distributed energy resources in a microgrid is carried out in [25] using DRL, where uncertainties in electricity consumption and renewable output are considered. DRL has been applied for the management of charging stations and electric vehicles as well. For example, decentralized scheduling solutions for multiple EV charging stations are performed in [26], and federated learning is incorporated with DRL in [20] to ensure consumer privacy for managing charging stations. Similarly, in [27], a deep neural network is used to predict traffic patterns and reinforcement learning to improve prediction accuracy. Similarly, a DRL-based EV charging control mechanism is proposed in [28] to enhance the consumption of renewables and enhance the SOC of EVs. Finally, minimization of charging time and origin-destination distance are considered for EVs in [29].

It can be observed from the existing literature that most of the studies have focused on the planning phase optimization of BESS for FCSs [12–15], i.e., optimal sizing of the BESS. However, the daily operation of BESS is also required to be optimized to maximize the benefit of BESS for the FCS and minimize the grid impacts of FCS. Most of the studies conducted on the daily operation of BESS have considered deterministic or stochastic models to operate FCSs including BESS. However, the model complexity of such mathematical problems increases with an increase in the number of random variables. Therefore, DRL can be a suitable method for the energy management of FCSs coupled with BESS due to the dynamic nature of the problem and several uncertainties associated with the EV demand modeling. In addition, the DRL model is immune to small perturbations thus making it suitable for the small uncertainties which may occur beyond the training data. Therefore, the trained model can be used for real-time management of FCSs with BESS to minimize the cost and impact of FCS on the power network under arrival and departure time uncertainties of EVs. Most of the existing studies on DRL-based methods [20,26] have only focused on maximizing the profit of FCS while ignoring the grid impacts of FCSs.

To address the shortcoming in the existing literature, mentioned in the previous paragraphs, a DRL-based operation model is developed in this study to operate FCSs with BESS. The state-of-the-art DRL method, soft actor-critic (SAC), is adopted in this study due to its superior traits over other DRL methods, such as fast convergence, stable output, and the ability to avoid local optima trapping. The developed model is trained using the data of one year considering local driving patterns and market price in Korea for the year 2020 to cover seasonality effects. The model trained using daily data is used due to faster convergence with similar results, as compared to the other two models (weekly and biweekly). The performance of the developed model is tested for operation during different day types (weekdays and holidays) and extreme uncertainty cases. The proposed method has successfully operated the FCS with BESS in the desired fashion under diverse conditions. The major contributions of this study are as follows.

- A state-of-the-art DRL method, the SAC method, is adopted to operate FCSs with BESS. A comprehensive reward function is devised to capture different uncertainties in the EV load and upstream grid price.
- Performance of different state-of-the-art DRL methods such as deep deterministic policy gradient (DDPG) and twin delayed DDPG (TD3) is analyzed and compared with SAC for the problem under consideration. In addition, the SAC model has been trained using three sets of data, i.e., daily, weekly, and bi-weekly.
- EV load is estimated considering different vehicle types (commercial and private) for different day types (weekdays and holidays) for an entire year.
- Extreme cases of the market price are considered to evaluate the robustness of the proposed method. In addition, the robustness of the proposed method is analyzed for arrival time lagging and leading cases of the EV fleet.

2. Fast charging stations and battery energy storage

Most EV owners prefer to recharge their EVs at home. Therefore, the peak load of EVs may coincide with the residential evening peak load. It became more severe in the case of clustered EVs, where several EVs are connected to a single charging station. Clustered EVs may overload the distribution transformer since it is the first bottleneck for enhanced EV penetration [30]. Therefore, BESS can be utilized to reduce the load of FCS during system peak hours and thus avoiding the overloading of transformer and power lines.

2.1. System configuration

An overview of the configuration of the proposed FCS with BESS is shown in Fig. 1. It can be observed that the total energy demand of EVs in the FCS (P_t^{EV}) can be fulfilled by either discharging the BESS (P_t^{B-}) and/or buying power from the grid (P_t^{Buy}). Similarly, the BESS can be charged (P_t^{B+}) by buying power from the grid. The power balance of the FCS at each time interval t can be determined by using Eq. (1), where the sum of power bought from the grid and power discharged from the BESS need to be equal to the EV fleet load and power charged to the BESS. The EV fleet load can be computed by summing loads of individual EVs ($P_{t,v}^{ev}$) as given by Eq. (2), where V refers to the total number of EVs in the fleet. It is worth noting that due to the presence of the BESS, the charging behavior of the EVs will not change significantly throughout the day. For example, EVs needing recharge during system peak hours will be able to charge using the energy stored in the BESS. However, from the power system perspective, it will reduce load since power is not bought from the grid during peak hours, i.e., peak shaving.

$$P_t^{Buy} + P_t^{B-} = P_t^{EV} + P_t^{B+} \tag{1}$$

$$P_t^{EV} = \sum_{v \in V} P_{t,v}^{ev} \tag{2}$$

The charging station operator (CSO) is responsible for managing the operation of the FCS with BESS. The CSO receives the market price signal information from the upstream grid and decides to charge/discharge the battery while fulfilling the energy needs of EVs. However, optimal operation of the FCS is challenging due to the involvement of uncertainty in the arrival and departure times of the EVs. In addition, the market price signals of the upstream grid are also uncertain. Therefore, the CSO will be equipped with the proposed DRL-based optimization algorithm to determine the optimal charging/discharging schedule of the BESS under these uncertainties. It will not only reduce the peak load of the FCS but also increase the revenue for the CSO by avoiding congestion/peak demand charges.

2.2. EV load estimation

To determine the optimal operation of the FCS, the net load of the EV fleet is required at each interval of the day. The load of EVs is uncertain and depends on several factors such as daily mileage of EVs, energy consumption per km, and state-of-charge (SOC) at the return time of EVs. The daily mileage of vehicles could be different based on the purpose of usage, i.e., private or commercial. Commercial vehicles tend to travel more during weekdays as compared to private vehicles. Therefore, the daily mileage of EVs is taken as a random variable with lognormal distribution ($f(d|\mu, \sigma)$). It can be mathematically represented as Eq. (3), where d is the daily mileage.

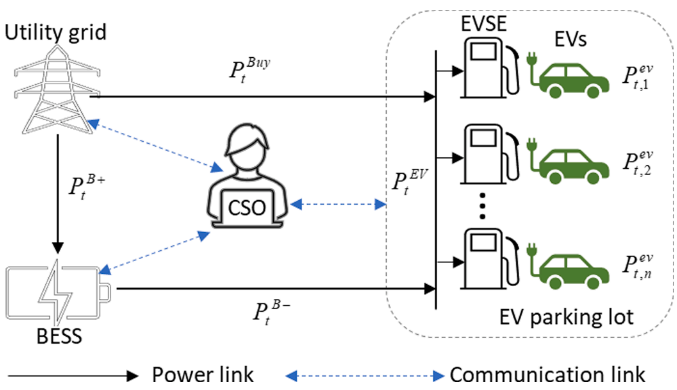

Fig. 1. Configuration of the proposed fast charging station with energy storage system.

$$f(d|\mu, \sigma) = \frac{1}{d.\sigma.\sqrt{2\pi}} e^{-\frac{(lnd-\mu)^2}{2\sigma^2}}, d > 0 \tag{3}$$

An overview of the daily mileage of private and commercial vehicles is shown in Fig. 2a. It has been demonstrated in several studies that the daily mileage of vehicles follows a lognormal distribution function [12, 31], i.e., the probability of occurrence of negative distance is zero and positive distance tail extends to infinity. Therefore, in this study also, the daily mileage is taken as a lognormal function. Six-year vehicle traveling data of South Korea [32] is used in this study, assuming that EVs will also follow a similar traveling pattern. Based on the six-year traveling data, the daily mean mileage for commercial vehicles turns out as 95.2 km and the standard deviation as 66.9 km. Similarly, the mean and standard deviation for private EVs turned out to be 34.1 km and 14.4 km, respectively. Vehicles are divided into four categories (sedan, vans, freight, and special) and four types of fuels (gasoline, diesel, LPG, and other fuel) are considered for each vehicle type. A total of 17 regions/cities are included in the database and the data are publicly available on the Korea Statistical Information Service (KOSIS) website [32].

It is assumed that SOC reduces linearly [31] and initial SOC at the arrival time (S^{ini}) can be computed using Eq. (4). $S^{max}$is the SOC at the beginning of the day while $\eta$is the energy consumption per km and D is the distance traveled by the EV.

$$S^{ini} = S^{max} - \eta \cdot D \tag{4}$$

The initial SOC is also taken as a random variable with a lognormal distribution. The SOC of private and commercial vehicles based on the mileage of EVs is shown in Fig. 2b. Due to the linear relationship between the daily mileage and SOC, the SOC also follows a lognormal distribution function. It can be observed that private EVs do not need a recharge on daily basis and the mean SOC is about 70% at the end of the first day. However, most of the private EVs will need a recharge after the second day, since the mean SOC drops below 50% after 2 days. However, commercial vehicles need to be recharged on daily basis (on working days), since the mean SOC drops below 50% at the end of each working day.

The return time of EVs is also taken as a random variable with a normal distribution ($h(t|\mu_t, \sigma_t)$), as shown in Eq. (5).

$$h(t|\mu_t, \sigma_t) = \frac{1}{\sigma_t.\sqrt{2\pi}} e^{-\frac{(t-\mu_t)^2}{2\sigma_t{}^2}} \tag{5}$$

The total load of EV fleet for any time t can be determined using Eq. (6), where P_v is the charging level of the EV and Δt is time step in hours.

$$P_t^{EV} = \Delta t.\sum_{v \in V} P_v.h(t, v).\vartheta\left(\mathrm{S}^{ini+t}, v\right) \tag{6}$$

The probability of power required to recharge an EV from the initial level ($\vartheta(\mathrm{S}^{ini+t}, v)$) is given by Eq. (7). V_v^{cap} is the capacity of the EV battery and $g(S^{ini} \leq S^t \leq S^{max}, v)$ is the probability of SOC being in the range of $S^{ini}$and S^{max}.

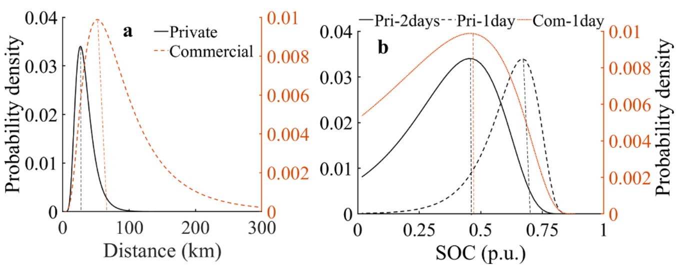

Fig. 2. Probability densities of private and commercial vehicles [12]; (a) daily mileage; (b) energy consumption.

$$\vartheta\left(S^{ini+t}, v\right) = \begin{cases} 1 \quad if\left(V_v^{cap} / P_v.\Delta t\right) - S^{ini} > 1 \\ g\left(S^{ini} \le S^t \le S^{max}, v\right) else \end{cases} \tag{7}$$

2.3. Battery energy storage model

In addition to the EV load model, the model of BESS is also required to realize different charging and discharging limits. The energy present in the BESS at any interval t is the sum of the initial energy (P_B^{Ini}) and the sum of power charged or discharged till interval t ($\sum_{\tau \le t}(\eta^+.P_\tau^{B+} - P_\tau^{B-} / \eta^-)$). η^+ and η^+represent the charging and discharging efficiencies, respectively. Eq. (8) implies that the SOC cannot exceed the upper limit of the SOC (SOC^{max}) and Eq. (9) implies that SOC cannot be below the lower SOC limit (SOC^{min}). In these equations, B^{Cap} is the capacity of the BESS in kWh.

$$P_B^{Ini} + \sum_{\tau \le t}\left(\eta^+.P_\tau^{B+} - P_\tau^{B-} / \eta^-\right) \le B^{Cap}.SOC^{max} / 100 \tag{8}$$

$$B^{Cap}.SOC^{min} / 100 \le P_B^{Ini} + \sum_{\tau \le t}\left(\eta^+.P_\tau^{B+} - P_\tau^{B-} / \eta^-\right) \tag{9}$$

Finally, (10) limits the charging (P^{Br+}) and discharging (P^{Br-}) rates of the BESS. It is worth mentioning that this model will not be used in its current form for DRL-based optimization. The constraints of this model will be realized in a simplified manner, which is discussed in the following section.

$$0 \le P_t^{B+} \le P^{Br+}, \quad 0 \le P_t^{B-} \le P^{Br-} \tag{10}$$

3. DRL-based operation of BESS

The arrival time of EVs at the charging station is uncertain and so are the power system load and market price signals. The estimated load, based on the probability distribution functions discussed in the previous section, for EVs on weekdays and holidays is shown in Fig. 3a. Similarly, the hourly load of the Korean power system on a typical summer weekday and holiday is shown in Fig. 3b. It can be observed that the EV peak load coincides with the system peak load. In addition, the fluctuations in load demands are also visible which translates to the market price signals. The optimal operation of BESS in FCSs is a multiperiod stochastic optimization problem. Operation of BESS cannot be determined with information of only the current intervals, it requires the SOC information of previous intervals and load information of upcoming intervals. This problem can be solved using stochastic optimization as well, but it has several limitations as compared to the DRL method, which are as follows.

- Being a multiperiod optimization problem, the complexity significantly increases with an increase in the number of intervals. The size of scenarios increases exponentially with an increase in interval length (365×48 in this study) and scenario reduction methods are required, which introduce further errors in the model.

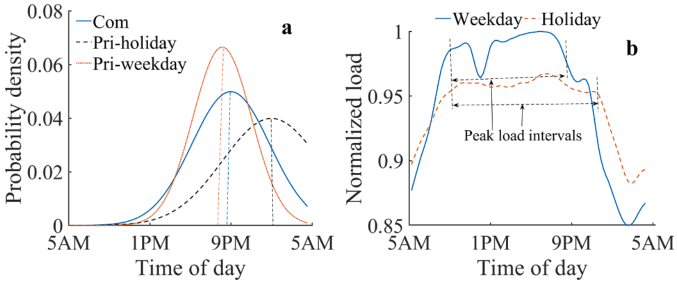

Fig. 3. a) EV load profiles for different day types [12]; b) normalized Korean power system load on a typical weekday and a holiday.

- Stochastic optimization requires complete knowledge of stochastic processes (EV behavior and market price), which is not possible. Therefore, PDFs are used to approximate the behavior, which results in the inaccuracy of the model.
- The stochastic problem becomes computationally intractable with a higher number of EVs since each EV introduces several random variables, such as arrival time, departure time, SOC, etc. Several scenarios need to be generated for each random variable.

Therefore, the DRL-based operation of FCS is proposed in this study, which does not require any model of the system and approximation of underlying uncertainties. Instead, it learns over time by interacting with the environment.

The basic idea of DRL is to combine a deep neural network with a reinforcement learning model (Q-model) to overcome the issues of Q-learning, i.e., the inability to deal with continuous state space or environments with uncertainties [21]. There are several variants of DRL but SAC is the state-of-the-art method and it has several advantages, such as fast learning, stable operation, and is immune to trapping in local optima [33]. Therefore, in this study, SAC is used for the optimal operation of the FCS. In SAC, the objective of conventional reinforcement learning is modified by adding an entropy term, which measures the predictability of the random variable [34]. A policy network is also used to avoid the trapping of the agent in the same action repeatedly. Two neural networks (actor and critic) are used together with the policy network and the operation principle with reference to the FCS optimal operation problem is discussed in the following sections.

3.1. SAC-based problem formulation

The flowchart of the proposed SAC-based method for optimal operation of FCS with BESS is shown in Fig. 4. The objective of this formulation is to determine the optimal charging/discharging schedule for the BESS under EV demand and price uncertainties. The operation algorithm based on SAC is divided into an actor and a critic. The critic updates the action-value function by evaluating the policy. The action-value is the measure of the net discounted reward and is then fed to the actor [33]. The actor directly interacts with the environment by choosing an action based on the current state. The action comprises of a number between [−1 1] in our case to depict the SOC level of the BESS, where positive values refer to discharging and negative values refer to charging. The state contains information about the current interval, SOC, market price, and EV fleet load. The step-by-step implementation process of the SAC is as follows.

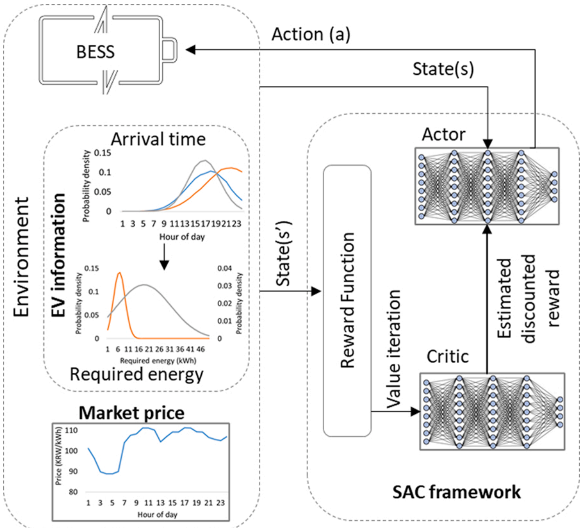

Fig. 4. Flowchart of FCS with BESS based on the SAC model.

1 Initialize five deep neural networks (a policy network, two Q networks, and two target networks).
2 Initialize an empty memory of size *D*.
3 Choose an action based on the current policy and receive an immediate reward then transition to a new state.
4 Store the information {*s, a, r, s*'} in memory *D*.
5 Select a random mini batch from the memory *D* and a target network then calculate the loss.
6 Update weights of all deep neural networks using gradient descent (Q-functions) and gradient ascent (policy) functions.
7 After a few iterations, update the target network using a soft copy from Q-functions.
8 Repeat steps 3 to 7 for a sufficient number of times (epochs).
9 Use the trained model to determine the optimal charging/discharging schedule of the BESS.

3.2. Reward function and operation principal

The reward function is the brain for any reinforcement algorithm, which dictates the actions of the agent. Therefore, a comprehensive reward function is devised to optimally operate the FCS with BESS under different uncertainties. The formulation of the reward function depends upon the objective of the problem under consideration and it needs to contain the decision parameters or factors related to the decision parameters. The objection of the problem under consideration in this study is to minimize the peak load of the charging station during system peak hours by optimally operating the BESS. The electricity price (market price) is a reflection of the power system load in any power system, i.e., higher during peak load periods and lower during off-peak periods. Therefore, in this study, the market price is used to determine the system peak load intervals. To incorporate all these parameters, a reward function (R^{ep}) is devised in this study, which is comprised of four reward parts as shown in Eq. (11).

$$R^{ep} = \sum_{t\in T}\left(\frac{r_t^{load} + r_t^{B+} + r_t^{B-} + r_t^{peak}}{D\cdot\Gamma}\right) \tag{11}$$

It is worth noting that multiple reward functions could give the same result, thus the reward function devised in this study is not the only reward function to get the same/similar results. The four factors of the reward function and rationale behind the formulation of each factor are as follows.

- **Load reward (r_t^{load}):** The objective of this factor is to minimize the difference between the threshold load level (P^{th}) and the net demand of the FCS with BESS ($P_t^{load} + P_t^{B+} - P_t^{B-}$), as given by Eq. (12). The threshold level is taken as the mean value of the charging station's yearly load. This factor dictates the BESS when to charge and when to discharge, i.e., charging during lower load periods and discharging during higher load intervals of the charging station will minimize the error (maximize the reward). However, this factor alone does not determine the final charging and discharging, other factors also play role in decisions regarding BESS charging/discharging.

$$r_t^{load} = -\left(\left(P_t^{load} + P_t^{B+} - P_t^{B-}\right) - P^{th}\right)^2 \tag{12}$$

- **Battery charging reward (r_t^{B+}):** The objective of this reward factor is to charge the BESS during low price intervals. However, BESS needs to know the price information of upcoming intervals as well to compare and decide whether the price of the current interval is higher or lower. To achieve this objective, the current interval's price (PR_t) is compared with a randomly selected price from the next *K* intervals (PR_K), as given by Eq. (13). The choice of next interval is randomly selected in each episode.

$$r_t^{B+} = P_t^{B+}\cdot(rand(PR_K) - PR_t)\mathrm{K} \in [t, \Gamma] \tag{13}$$

- **Battery discharging reward (r_t^{B-}):** The objective of this reward factor is to discharge the BESS during peak price intervals. Similar to the charging case, BESS needs to know the price of upcoming intervals; therefore, this factor is realized in a similar way to that of charging reward, as given by Eq. (14).

$$r_t^{B-} = P_t^{B-}\cdot(PR_t - rand(PR_K))\mathrm{K} \in [t, \Gamma] \tag{14}$$

- **Peak load reward (r_t^{peak}):** The objective of this factor is to penalize deviations from the threshold level during peak load intervals of the power system. System peak load intervals are identified via peak beginning interval (t^{pb}) and peak ending interval (t^{pe}). It can be observed from Eq. (15) that this factor is zero for the off-peak load intervals. It can also be observed that the penalty will be double if the power deviates from the threshold during system peak load intervals, i.e., once due to the first factor and once due to the fourth factor. Therefore, the model will try to remain close to the threshold during system peak intervals, i.e., buying from the grid will be minimized.

$$r_t^{peak} = \begin{cases} r_t^{load} if t \in \left[t^{pb}, t^{pe}\right] \\ 0 else \end{cases} \tag{15}$$

Finally, the total reward function is normalized with the number of days (*D*) and intervals (Γ) per day to analyze different training horizons, Eq. (1). All these four factors collectively determine the charging/discharging behavior of the BESS to flatten the load profile of the charging station and reduce buying power from the grid during system peak load intervals.

The SOC of the BESS is updated after each step based on the action taken by the agent, as given by Eq. (16). ΔSoC_{t+1} is the action taken by the agent in the range of [−1,1].

$$SoC_{t+1} = SoC_t + \Delta SoC_{t+1}, \Delta SoC_{t+1} \in [-1, 1] \tag{16}$$

Based on the current and previous SOC levels along with the capacity of the BESS (B^{Cap}), battery power (P_t^B) can be determined using Eq. (17).

$$P_t^B = B^{Cap}\cdot(SoC_{t-1} - \Delta SoC_t) \tag{17}$$

After analyzing the sign of the battery power, charging (negative) and discharging (positive) amounts can be determined using Eq. (18).

$$\left.\begin{array}{c} if P_t^B < 0 P_t^{B+} \\ else P_t^{B-} \end{array}\right\} = P_t^B \tag{18}$$

The charging and discharging power amounts determined via this method are used in the reward function, as discussed in the previous section.

4. Numerical simulations

In this section, the performance of the proposed method is evaluated for a multi-unit residential apartment, where parking space is shared by the residents. The charging station has a BESS unit and is also connected with the grid, i.e., can buy power from the grid when required. The residential electricity tariff will be applicable for the charging station as well since it is in a residential building. It is worth noting that the proposed model is tested for a residential apartment complex having shared parking space. However, the proposed method can be used for any type of building (having sharing station and on-site energy storage system) if the data of that building is available. The developed mathematical models are generalized and can be used for any type of building.

4.1. *Input data*

EV load profiles are obtained for a fleet of 200EVs for each day of the year 2020 using the EV load estimation method presented in the previous section. Similar to Hussain et al. [12], 5% of the fleet is considered as commercial EVs while 95% as private EVs, and the operation range of battery SOC in each EV is set to [0.1–0.9]. The commercial EVs are considered only on weekdays while private EVs are recharged on alternate days, i.e., 50% of private EVs are recharged on each day. During each interval, the load of each EV is computed considering the arrival probability density functions. Then, the load of all EVs is accumulated to obtain the total charging station load (EV fleet) for each interval of the day. The load of the EV fleet on a selected weekday and a holiday is shown in Fig. 5. The statistical data of EV arrival time at a residential apartment complex is shown in Table 1, which is based on the traveling pattern of vehicles in Korea for commuting and business purposes [35]. Similarly, the statistical parameters of daily mileage are also derived from six years of traveling data of Korea [32]. The market price for the year 2020 in Korea [36] is analyzed and upper and lower bounds are defined for different periods of the day, different for weekdays and holidays. Then an uncertainty factor is added to each price level to mimic the deviations during different days. For the sake of visualization, an overview of the hourly market price on a selected weekday and holiday is presented in Fig. 5. However, the model is trained using data (market price and EV fleet load) of a whole year. Therefore, it can be used for optimizing the operation of the charging station for any season/day of the year. The size of the BESS in the FCS is taken as 400 kWh and the SOC is set to operate between 0.1 and 0.9 to reduce the degradation by avoiding deep discharging and overcharging. The roundtrip efficiency of the BESS is taken as 90%.

4.2. *Comparison of DRL methods*

In this section, the performance of three of the most advanced DRL methods (DDPG, TD3, and SAC) is compared and analyzed. All the models are trained using one-day data, which contains 48 samples and each sample corresponds to 30 min. The per-episode reward and moving average reward of all three methods are shown in Fig. 6. It can be observed that SAC outperforms both TD3 and DDPG in terms of convergence speed. For example, SAC has converged under 500 episodes while TD3 and DDPG took over 1000 episodes. This is because the SAC learns stochastic policies via a maximum entropy objective [37]. In addition, it can also be observed that SAC converges to a higher reward as compared to TD3 and DDPG and it is in alignment with other studies [23,37]. Finally, the fluctuations in SAC are lower as compared to the other two methods, i.e., the stability of SAC is higher. This is also due to the learning of a stochastic policy while maximizing entropy as explained by the founders of SAC [37]. Due to these desirable traits of SAC over other state-of-the-art DRL methods, SAC is used in this study and is discussed in the following sections.

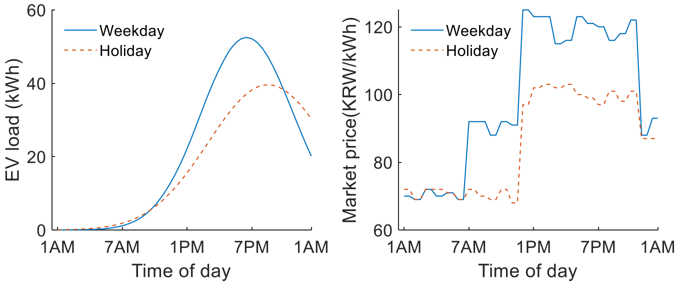

Fig. 5. EV load and market price on a selected weekday and a holiday.

Table 1
Statistic parameters of daily vehicle arrival time and mileage.

Parameter		Mean	Standard deviation	Unit
Arrival time	Private EVs on weekdays	19	4	hour
	Private EVs on holidays	21	5	hour
	Commercial EVs	18	3	hour
Daily mileage	Private EVs	34.1	14.4	km
	Commercial EVs	95.2	66.9	km

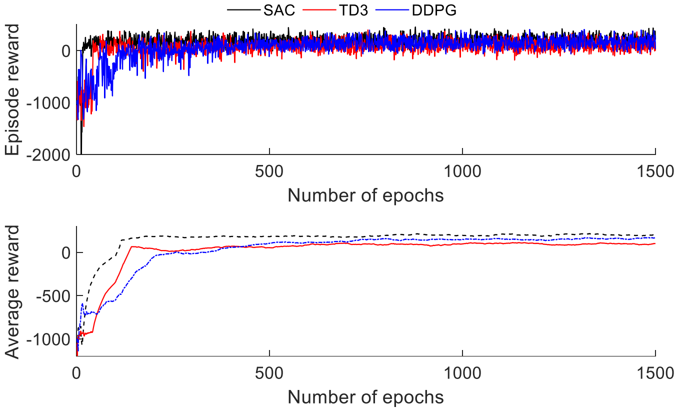

Fig. 6. Convergence analysis of different DRL methods.

4.3. *Model training results*

4.3.1. *Hyperparameter tuning for SAC*

Hyperparameter tuning is one of the tricky tasks in most DRL methods. However, SAC has the advantage over other DRL methods since only one hyperparameter needs to be tuned in the case of SAC [37]. The temperature hyperparameter needs to be tuned since the performance of the maximum entropy reinforcement learning depends on the scaling factor and it has to be compensated by the choice of suitable temperature [38]. The authors of SAC have proposed an auto-tuning framework for SAC in [38] where an automatic gradient-based temperature tuning method is proposed. The proposed method adjusts the expected entropy over the visited states to match a target value. It has been tested for several cases and found that it largely eliminates the need for per-task hyperparameter tuning [38]. A recent study has proposed a method (namely Meta-SAC), where metagradient is used along with a novel meta objective to automatically tune the entropy temperature in SAC [39]. These methods can be used to auto-tune the temperature hyperparameter in SAC. In addition, the number of hidden layers and the number of neurons in each hidden layer also impact the performance of the method. Generally, the number of hidden layers is increased if the performance of the method is not up to the mark but it increases training time. Therefore, a trade-off need to be selected. The hyperparameters

Table 2
Hyperparameters used for simulating the SAC model.

Parameter	Value	Parameter	Valve
Optimizer	Adam	Number of samples per minibatch	256
Learning rate (actor)	2.5×10^{-4}	Nonlinearity	ReLU
Learning rate (critic)	2.5×10^{-3}	Target smoothing coefficient (τ)	0.005
Discount factor (γ)	0.999	Target update interval	1
Replay buffer size	10^{6}	Gradient steps	1
Number of hidden layers (all networks)	2	Input dimension	4
Number of hidden units per layer	256	Reward scale	$D{\cdot}\Gamma$
Action dimensions	1	Entropy target	-1

used for simulation of the proposed network are summarized in Table 2.

4.3.2. Performance analysis of SAC

To evaluate the performance of the proposed SAC-based operation of FCS with ESS, three different lengths of data are used for training the model. The training data for the first case (daily) contains 48 samples and each sample corresponds to 30 min. The weekly case contains 7 × 48 and the bi-weekly case contains 14×48 samples. However, the reward is computed using reward function (11), which gives normalized rewards. An overview of the convergence of all three cases is presented in Fig. 7. It can be observed that there is no significant difference among the average rewards obtained in all three cases. The magnitude of reward fluctuations decreases with an increase in the training data size due to the averaging effect. However, larger data takes more epochs for convergences, as can be observed from Fig. 7. In addition, the time required per epoch also significantly increases with an increase in the data size. The operation results of a selected week for all three cases are shown in Fig. 8, where the same price and EV load are used for each case. The developed algorithm successfully uses the battery in the set operation range (10% – 90%) for all the classes during all the days of the week. It can be observed that the overall battery SOC profile is the same for all cases. In some cases, the battery is charged to a higher level during off-peak intervals, during some days. However, the discharging level is also higher during those cases/days. Therefore, the energy utilized is the same for all cases. Due to this similarity in results, a one-day training model is used for the rest of the analysis due to its faster convergence while giving the same results with larger data sizes.

4.4. Peak shaving impact analysis

To analyze the peak shaving impact of the proposed SAC-based operation algorithm, two days (a weekday and a holiday) are selected in this section. It can be observed from Fig. 9 that BESS is charged during the off-peak (1 AM–6 AM) and shoulder peak (7 AM–10 AM) intervals. Therefore, the FCS load has increased during those intervals. Similarly, BESS is discharged during peak intervals to reduce the peak load of the FCS. The original peak was around 7 PM with a magnitude of 54 kW, which is reduced to 29 kW by the proposed method. The reduction in peak load is about 46%. Similarly, the peak shaving impact on a selected holiday is shown in Fig. 10. The BESS charging takes place during off-peak intervals and is discharged during peak intervals, similar to the

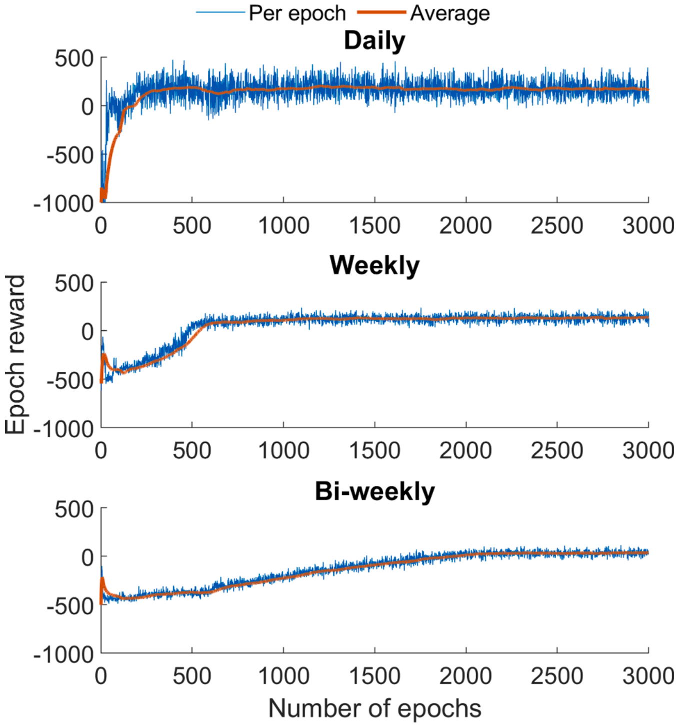

Fig. 7. Convergence analysis of SAC under different training data lengths.

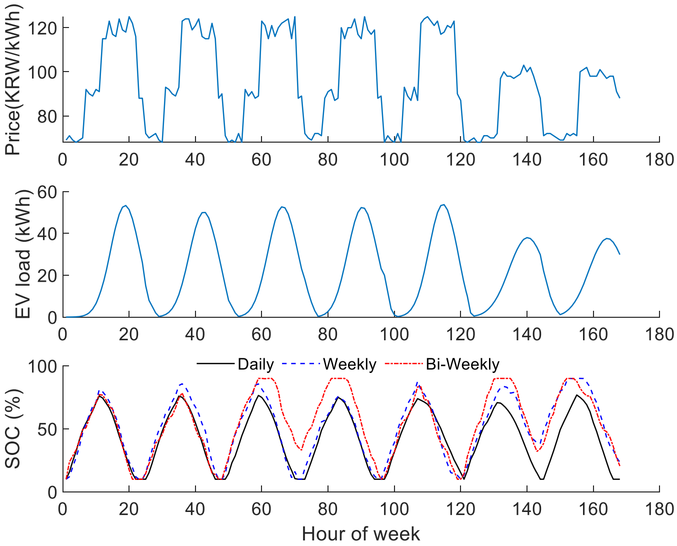

Fig. 8. Weekly FCS operation results under different load and market prices.

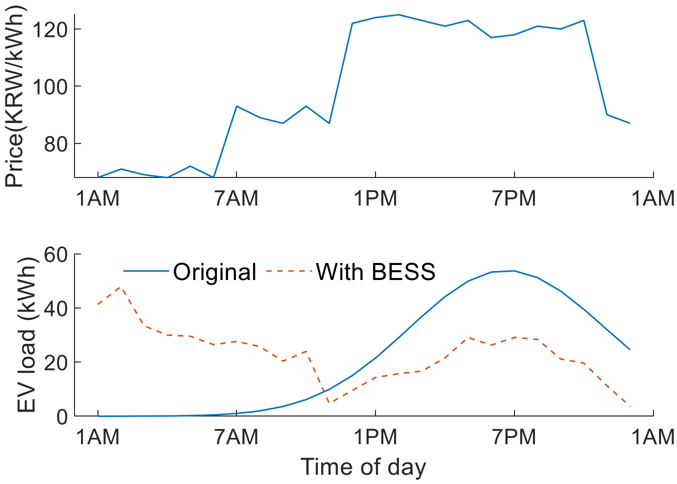

Fig. 9. Peak shaving results of a selected weekday.

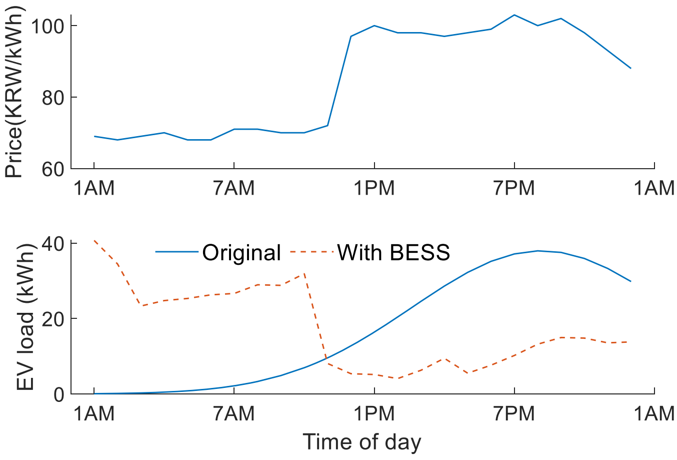

Fig. 10. Peak shaving results of a selected holiday.

weekday. The original peak was at 8 PM with a magnitude of 38 kW which is reduced to 13 kW by the proposed algorithm. The reduction in peak load is about 65% in this case. Higher peak shavings during holidays are expected since the FCS load is lower during holidays due to the absence of the commercial EV load. The peak load of the FCS after using the proposed method has been reduced to 29 kW (7 PM) and 15 kW (9 PM) for weekdays and holidays, respectively. It can be concluded that the proposed method operates the BESS in a desired manner to reduce FCS peak load during both working days and holidays.

4.5. Robustness analysis

There are several uncertainties associated with the BESS management in the FCS. All these uncertainties either result in a change in the load profile of the PEV fleet or the market price of the utility grid, applied to the FCS. Therefore, the robustness of the proposed method is analyzed for some extreme cases for both uncertain parameters (load and market price) in this section. In the first case, the load profile of the EV fleet in the FCS is time-advanced by 1 and 2 h (lagging) and time-delayed by 1 and 2 h (leading). Including the normal scenario, five scenarios are developed as shown in Fig. 11. The results obtained for all these five scenarios are shown in Fig. 12. It can be observed the charging behavior is similar for all five scenarios, i.e., reach the maximum charging level around 12 PM. However, the discharging behavior significantly changes for each scenario. For example, for the time-advanced scenarios, more energy is discharged earlier due to the early arrival of peak load. Similarly, for the time-delayed scenarios, more energy is discharged in later inter intervals due to the shift of peak load towards the later intervals. It can be concluded that the trained SAC-based operation model is robust enough to operate in the desired fashion under different uncertainties in the arrival time of EVs, i.e., shifted load profiles of FCS.

In the second case, the gap between the off-peak and the peak price is varied and three scenarios are simulated for a weekday. In the first scenario, the gap between the off-peak and peak price is the highest while in the second scenario the gap is the lowest. Finally, in the third scenario, the gap is randomly generated between the upper and lower bounds. An overview of the three price scenarios is presented in Fig. 13. The interval-wise peak load shaving results under these three scenarios are shown in Table 3. It can be observed that peak load is shaved for different intervals to minimize the net load of the FCS. The last column shows the average load shed for all the intervals. It can be observed that the highest peak shaving was for the case of the highest price gap and the lowest for the lowest price gap. This is due to a higher opportunity for savings for the FCS operator during the highest gap scenario. In addition, the market price is an indication of the load level in the power system. Therefore, it is desired to reduce load during higher peak price days, which was successfully achieved by the developed method. Finally, the average peak shaving for the random case was in between the two cases, which is also expected. The price, in this case, is in between the two extreme cases. It can be concluded that the proposed method can reduce the peak load of the FCS in accordance with the power system congestion conditions, i.e., different market price scenarios.

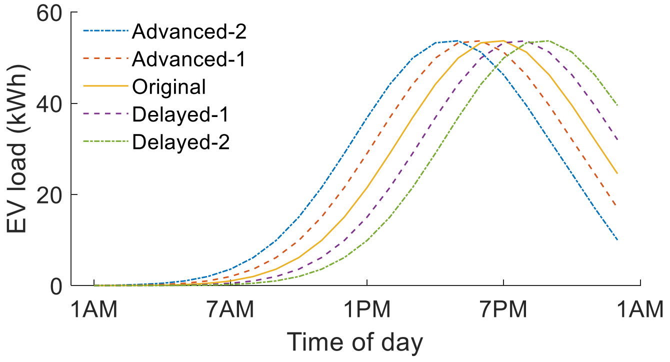

Fig. 11. EV fleet profile leading and lagging scenarios in time.

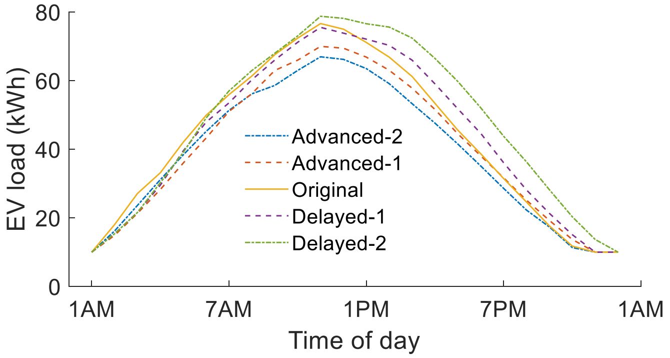

Fig. 12. SOC profiles of BESS under time leading and lagging scenarios.

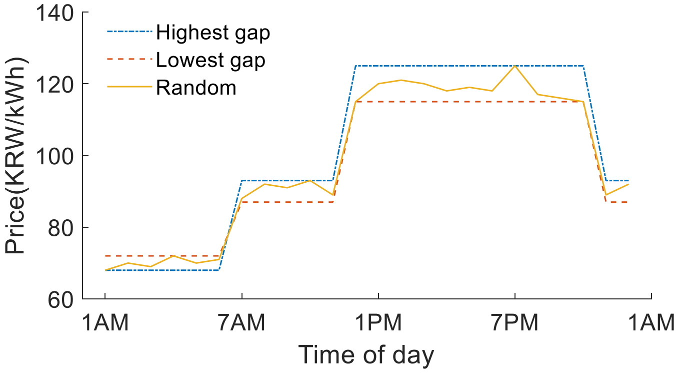

Fig. 13. Market price scenarios.

Table 3
Load reduction during peak intervals under different price gaps.

Interval	Highest gap	Lowest gap	Random
11AM	69.1%	72.0%	45.9%
12PM	55.6%	49.1%	59.5%
1PM	63.4%	50.5%	47.6%
2PM	53.7%	48.9%	58.5%
3PM	48.4%	49.4%	47.8%
4PM	52.0%	52.2%	60.1%
5PM	51.7%	51.1%	58.9%
6PM	55.9%	52.7%	56.5%
7PM	60.8%	48.6%	50.8%
8PM	51.8%	52.6%	62.0%
9PM	56.8%	61.9%	58.4%
10PM	54.5%	63.1%	57.7%
Average	56.2%	54.3%	55.3%

5. Conclusions

A deep reinforcement learning-based model is developed for optimizing the operation of fast charging stations equipped with an energy storage system. Among different deep reinforcement learning methods, the soft actor-critic method is adopted due to its superior performance in terms of convergence speed and optimality. The convergence speed of soft actor-critic was over 2-times faster than those other deep reinforcement learning methods such as deep deterministic policy gradient and twin delayed deep deterministic policy gradient. The proposed model has been trained offline using data of one year to cover seasonality and the impact of different working days and holidays on the load of the fast-charging station. A comprehensive reward function is developed to train the model under diverse uncertainties, both in electric vehicle load and upstream grid price. The trained model can be used for the real-time operation of a fast charging station with a battery energy storage system under load and price uncertainties. Simulation results have shown that the proposed method can optimize the operation of the battery during both weekdays and holidays to minimize the peak load of the fast-charging station. Peak shaving of up to 46% was observed for weekdays and up to 65% was observed for holidays. It has been demonstrated that the proposed method can operate the battery energy storage system under different uncertainty scenarios in the arrival time of electric vehicles, i.e., advanced and delayed arrivals. Similarly, the performance of the proposed method under extreme cases of market price uncertainties is also demonstrated through simulations. Simulation has shown that up to 56% of peak shaving can be achieved for weekdays with the highest difference in peak and off-peak prices.

The proposed method has been tested for a single charging station, in a residential apartment, having several electric vehicles. Analysis of the proposed method for different building types such as commercial, industrial, and mixed building types will be a valuable extension of this study. In addition, the model has been trained and tested using 30 min resolution data of charging station load and market price signals. However, using further granulated data (1 – 15 min) could be more beneficial for load estimation of the fast charging station.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Incheon National University Research Grant in 2021.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.epsr.2022.108087.

References

[1] M.A.H. Rafi, J. Bauman, A comprehensive review of DC fast-charging stations with energy storage: architectures, power converters, and analysis, IEEE Trans. Transp. Electrif. 7 (2) (2021) 345–368.

[2] R. Kochhan, S. Fuchs, B. Reuter, P. Burda, S. Matz, M. Lienkamp, An overview of costs for vehicle components, fuels, greenhouse gas emissions and total cost of ownership update 2017, Tum Creat. (2017) 1–26.

[3] C. Shao, T. Qian, Y. Wang, X. Wang, Coordinated planning of extreme fast charging stations and power distribution networks considering on-site storage, IEEE Trans. Intell. Transp. Syst. 22 (1) (2021) 493–504.

[4] K.L. Lopez, C. Gagne, M.A. Gardner, Demand-side management using deep learning for smart charging of electric vehicles, IEEE Trans. Smart Grid 10 (3) (2019) 2683–2691.

[5] S. Sachan, S. Deb, S.N. Singh, Different charging infrastructures along with smart charging strategies for electric vehicles, Sustain. Cities Soc. 60 (2020), 102238. December 2019.

[6] S.M.B. Sadati, J. Moshtagh, M. Shafie-khah, J.P.S. Catalão, Smart distribution system operational scheduling considering electric vehicle parking lot and demand response programs, Electr. Power Syst. Res. 160 (2018) 404–418.

[7] D. Sbordone, I. Bertini, B. Di Pietra, M.C. Falvo, A. Genovese, L. Martirano, EV fast charging stations and energy storage technologies: a real implementation in the smart micro grid paradigm, Electr. Power Syst. Res. 120 (2015) 96–108.

[8] S. Funke, P. Jochem, S. Ried, T. Gnann, Fast charging stations with stationary batteries: a techno-economic comparison of fast charging along highways and in cities, Transp. Res. Procedia 48 (2019) (2020) 3832–3849.

[9] A. Hussain, P. Musilek, Resilience enhancement strategies for and through electric vehicles, Sustain. Cities Soc. (2022). Jan.

[10] X. Duan, Z. Hu, Y. Song, Bidding strategies in energy and reserve markets for an aggregator of multiple EV fast charging stations with battery storage, IEEE Trans. Intell. Transp. Syst. 22 (1) (2021) 471–482.

[11] S. Negarestani, M. Fotuhi-Firuzabad, M. Rastegar, A. Rajabi-Ghahnavieh, Optimal sizing of storage system in a fast charging station for plug-in hybrid electric vehicles, IEEE Trans. Transp. Electrif. 2 (4) (2016) 443–453. Dec.

[12] A. Hussain, V.H. Bui, H.M. Kim, Optimal sizing of battery energy storage system in a fast EV charging station considering power outages, IEEE Trans. Transp. Electrif. 6 (2) (2020) 453–463. Jun.

[13] B. Sun, A multi-objective optimization model for fast electric vehicle charging stations with wind, PV power and energy storage, J. Clean. Prod. 288 (2021), 125564.

[14] A. Pal, A. Bhattacharya, A.K. Chakraborty, Placement of public fast-charging station and solar distributed generation with battery energy storage in distribution network considering uncertainties and traffic congestion, J. Energy Storage 41 (2021), 102939. April.

[15] J.A. Domínguez-Navarro, R. Dufo-López, J.M. Yusta-Loyo, J.S. Artal-Sevil, J. L. Bernal-Agustín, Design of an electric vehicle fast-charging station with integration of renewable energy and storage systems, Int. J. Electr. Power Energy Syst. 105 (2018) 46–58. March2019.

[16] S. Bahrami, Y.C. Chen, V.W.S. Wong, Deep reinforcement learning for demand response in distribution networks, IEEE Trans. Smart Grid 12 (2) (2021) 1496–1506, https://doi.org/10.1109/TSG.2020.3037066. Mar.

[17] T. Yang, L. Zhao, W. Li, A.Y. Zomaya, Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning, Energy 235 (2021), 121377, https://doi.org/10.1016/j.energy.2021.121377. Nov.

[18] S. Jaimungal, Reinforcement learning and stochastic optimisation, Financ. Stoch. 26 (1) (2022) 103–129, https://doi.org/10.1007/S00780-021-00467-2/FIGURES/8. Jan.

[19] H.M. Abdullah, A. Gastli, L. Ben-Brahim, Reinforcement learning based EV charging management systems-a review, IEEE Access 9 (2021) 41506–41531.

[20] S. Lee, D.H. Choi, Dynamic pricing and energy management for profit maximization in multiple smart electric vehicle charging stations: a privacy-preserving deep reinforcement learning approach, Appl. Energy 304 (2021), 117754, https://doi.org/10.1016/j.apenergy.2021.117754. Dec.

[21] V.-.H. Bui, A. Hussain, H.-.M. Kim, Double deep Q -learning-based distributed operation of battery energy storage system considering uncertainties, IEEE Trans. Smart Grid 11 (1) (2020).

[22] P. Kou, D. Liang, C. Wang, Z. Wu, L. Gao, Safe deep reinforcement learning-based constrained optimal control scheme for active distribution networks, Appl. Energy 264 (2020), 114772, https://doi.org/10.1016/j.apenergy.2020.114772. Apr.

[23] V.H. Bui, W. Su, Real-time operation of distribution network: a deep reinforcement learning-based reconfiguration approach, Sustain. Energy Technol. Assess. 50 (2022), 101841, https://doi.org/10.1016/j.seta.2021.101841. Mar.

[24] Y. Shang, et al., Stochastic dispatch of energy storage in microgrids: an augmented reinforcement learning approach, Appl. Energy 261 (2020), 114423, https://doi.org/10.1016/j.apenergy.2019.114423. Mar.

[25] L. Lei, Y. Tan, G. Dahlenburg, W. Xiang, K. Zheng, Dynamic energy dispatch based on deep reinforcement learning in IoT-driven smart isolated microgrids, IEEE Internet Things J. 8 (10) (2021) 7938–7953, https://doi.org/10.1109/JIOT.2020.3042007. May.

[26] M. Shin, D.H. Choi, J. Kim, Cooperative management for PV/ESS-enabled electric vehicle charging stations: a multiagent deep reinforcement learning approach, IEEE Trans. Ind. Inform. 16 (5) (2020) 3493–3503, https://doi.org/10.1109/TII.2019.2944183. May.

[27] T. Fu, C. Wang, N. Cheng, Deep-learning-based joint optimization of renewable energy storage and routing in vehicular energy network, IEEE Internet Things J. 7 (7) (2020) 6229–6241.

[28] M. Dorokhova, Y. Martinson, C. Ballif, N. Wyrsch, Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation, Appl. Energy 301 (2021), 117504. January.

[29] C. Zhang, Y. Liu, F. Wu, B. Tang, W. Fan, Effective charging planning based on deep reinforcement learning for electric vehicles, IEEE Trans. Intell. Transp. Syst. 22 (1) (2021) 542–554, https://doi.org/10.1109/TITS.2020.3002271. Jan.

[30] "Electric Vehicles: An Exploration on Adoption and Impacts | C4NET." https://c4net.com.au/projects/electric-vehicles-an-exploration-on-adoption-and-impacts/ (accessed Nov. 10, 2021).

[31] P. Zhang, K. Qian, C. Zhou, B.G. Stewart, D.M. Hepburn, A methodology for optimization of power systems demand due to electric vehicle charging load, IEEE Trans. Power Syst. 27 (3) (2012) 1628–1636, https://doi.org/10.1109/TPWRS.2012.2186595.

[32] "Statistical Database | KOSIS KOrean Statistical Information Service." https://kosis.kr/eng/statisticsList /statisticsListIndex.do?enuId=M_01_01&vwcd=MT_ETITLE&parmTabId=M_01_01&statId=1975011&themaId=#M2_3.2 (accessed Nov. 10, 2021).

[33] E. Anderlini, S. Husain, G.G. Parker, M. Abusara, G. Thomas, Towards real-time reinforcement learning control of a wave energy converter, J. Mar. Sci. Eng 8 (11) (2020) 845, https://doi.org/10.3390/JMSE8110845, 2020, Vol. 8, Page 845Oct.

[34] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018.

[35] J. Choi, W. Do Lee, W.H. Park, C. Kim, K. Choi, C.H. Joh, Analyzing changes in travel behavior in time and space using household travel surveys in Seoul Metropolitan Area over eight years, Travel Behav. Soc. 1 (1) (2014) 3–14, https://doi.org/10.1016/J.TBS.2013.10.003. Jan.

[36] "Electric Power Statistics Information System (EPSIS)." http://epsis.kpx.or.kr/epsisnew/selectMain.do?locale=eng (accessed Nov. 10, 2021).

[37] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor, PMLR (2018)

1861–1870. Jul. 03Accessed: Feb. 13, 2022. [Online]. Available, https://proceedings.mlr.press/v80/haarnoja18b.html.

[38] T. Haarnoja et al., "Soft actor-critic algorithms and applications," Dec. 2018, Accessed: Feb. 13, 2022. [Online]. Available: https://arxiv.org/abs/1812.05905v2.

[39] Y. Wang and T. Ni, "Meta-SAC: auto-tune the entropy temperature of soft actor-critic via metagradient," Jul. 2020, Accessed: Feb. 13, 2022. [Online]. Available: https://arxiv.org/abs/2007.01932v2.