

Online battery scheduling for grid-connected photo-voltaic systems[☆]

Samrat Nath^{*}, Jingxian Wu

Department of Electrical Engineering, University of Arkansas, Fayetteville, AR 72701, USA



ARTICLE INFO

Keywords:

Battery energy storage system (BESS)
Markov decision process (MDP)
Optimization
Photo-voltaic (PV)
Scheduling
Time-of-use (TOU)

ABSTRACT

We study the problem of determining optimum policy for managing battery energy storage system (BESS) in grid-connected photo-voltaic (PV) systems, where the stochastic electricity demands from the load are met from three sources: grid, PV energy, and BESS. BESS is used either to store excess energy generated from PV systems for later use, or to purchase energy from the grid when the time-of-use (TOU) pricing is lower. The objective is to identify the optimum charging/discharging schedule of BESS so that the long-term cost of energy purchased from the grid is minimized. The stochastic variabilities in loads and PV energy are captured by employing probabilistic models of periodic stochastic process with parameters estimated using historical data. The optimization problem is formulated under the framework of periodic discounted Markov decision process (MDP), and the problem formulation includes the aging effects of batteries and solar panels. The online optimization problem is solved by adopting a policy iteration approach tailored for periodic MDP. The proposed online scheduling algorithm provides periodic policies for a period of 24-hour, where the system model is updated every day based on load and PV energy from the previous day in a rolling horizon fashion. Simulation results demonstrate that the proposed algorithm can achieve a 41.6% reduction in annual utility bills compared to conventional systems without PV and BESS, thus ascertaining the values of installing BESS and PV systems.

1. Introduction

Grid-connected renewable energy systems, such as systems integrated with photo-voltaic (PV) and wind energy, are promising solutions to meet the increasing electricity demands while reducing greenhouse gas emissions and dependence on fossil fuels. Particularly, PV energy is gaining more and more popularity among residential and commercial users. With relatively low production and maintenance cost, the installation of a PV system is expected to achieve significant savings in energy cost. However, renewable energy sources like PV energy are non-dispatchable sources that are both intermittent and irregular in nature. These intermittent variations can pose significant challenges in the operation of the power grid at large penetration levels. However, the stochastic nature of renewable energy can be compensated by integrating the PV system with battery energy storage system (BESS) [1]. Energy storage technologies can address the challenges imposed by intermittent variations of PV sources by decoupling the time of energy generation and energy consumption, thus balance the energy demands and supplies at different time periods.

Battery management strategies in grid-connected renewable energy systems have been widely investigated in the literature with different

design objectives [1–13]. In [2], a predictive control system designed with dynamic programming (DP) is proposed to perform peak shaving by optimizing power flow management. Genetic algorithm (GA) is used to develop BESS management policies in both [3] and [4]. The objectives of [3] is to minimize the line loss of distribution systems, and that of [4] is to minimize the household energy cost. In [5], optimal energy management is performed for a grid-connected microgrid with a high degree of uncertainties, where a scenario-based technique is proposed in order to model the uncertainties in the output of PV generation, the load demand forecasting error, and the market price of electricity. The optimum design of BESS-assisted PV system is studied in Li and Wu [6], where the size of PV panels, the capacity of BESS, and the BESS charging/discharging scheduling are determined with the objective of minimizing the long-term average cost, including both energy cost and system cost. The optimization in Li and Wu [6] is performed in an offline fashion with complete knowledge of PV energy and load profile of an entire year. However, in order to capture the stochastic nature of both PV energy and load profile and to include the sequential characteristics of the decision problem, an online optimization approach is needed to provide optimum energy management in real-time based only on causal information of load profile and PV energy.

[☆] This work was supported by the U.S. National Science Foundation (NSF) under grant ECCS-1711087.

^{*} Corresponding author.

E-mail addresses: snath@uark.edu (S. Nath), wuj@uark.edu (J. Wu).

Online optimization of energy system can be performed by using stochastic dynamic programming approaches such as Markov decision process (MDP) [14]. MDP is a method for modeling systems with both probabilistic and non-deterministic behaviors, and it is widely used for solving scheduling problems [1,7,12,13,15,16]. The problem of optimum energy storage management and sizing is formulated as a stochastic dynamic program in Harsha and Dahleh [1], which aims at minimizing the long-term average cost of electricity and storage investment. It is shown that the optimum storage management policy has a dual-threshold structure under mild assumptions. Approximate dynamic programming (ADP) is adopted in Kwon et al. [7] to reduce the computational complexity of BESS scheduling in large-scale energy systems. In [12], a control policy using fitted Q-iteration algorithm is proposed for BESS scheduling in order to maximize self-consumption of local PV productions in a micro-grid. Periodic MDP is considered in Hu and Defourny [13], where periodic policies using value-iteration algorithm are proposed for grid-level storage management problem.

One of the important factors that can affect the scheduling performance of the system is the age of the devices [6]. Most of the above-mentioned works do not consider the aging effect of the devices, where the capacities and/or efficiencies of both BESS and PV system degrade gradually over time. Moreover, very few works consider the online (real-time) battery charge scheduling problem that takes account of the dynamic real-time system model. An online algorithm can provide adaptive solution based on the dynamic model as opposed to the fixed solution in offline optimization problems. To the best of our knowledge, there is no work in the literature that investigates the problem of online optimum battery charge scheduling in a grid-connected PV system while considering the aging effects of the batteries and PV panels.

In this paper, we study the optimum scheduling of BESS in grid-connected PV systems to minimize the average long term energy cost. BESS is used for two purposes: 1) to store excessive energy generated from PV systems for later use, and 2) to store energy purchased from the grid when the time-of-use (TOU) pricing is lower. The BESS can bridge the gap between energy supply and demands by shifting both PV energy and grid energy in time. The objective of the optimum design is to identify the optimum BESS charging/discharging policy such that the long-term cost of electricity purchased from grid is minimized. The key contributions of this paper can be listed as below:

- In order to capture both the deterministic and stochastic variability in the load and PV energy, probabilistic models for periodic stochastic process are employed, where the model parameters are estimated using historical data.
- The problem is formulated as a periodic discounted MDP, with each period corresponding to a period of 24 hours to account for the semi-periodic nature of loads and PV energy on a daily basis. The problem formulation considers the aging effects of batteries and PV panels.
- The problem is solved by developing a policy iteration algorithm that is tailored specifically for periodic discounted MDP. The algorithm can dynamically adjust the scheduling policy based on load and PV energy from the previous days in a rolling horizon fashion.

The remainder of this paper is organized as follows. Section 2 describes the system model. The optimization problem is formulated using the framework of periodic MDP in Section 3. The online BESS scheduling algorithm is developed in Section 4. Simulation results are given in Section 5, and Section 6 concludes this paper. A list of the notations used in this paper is summarized in Table 1.

2. System model

The system model is shown in Fig. 1. The system is assumed to operate in slotted time under hourly discretization, where $t \in \{1, 2, \dots\}$ denotes the time index.

Table 1

List of notations.

t	time index for hourly observations
$h(m)$	index of hours (months)
$\mathcal{H}(M)$	set of hours (months)
$q_t^{pv}(q_t^{ld})$	PV energy (electricity load) during time slot t (kWh)
q_t^b	charging/ discharging rate of battery (kW)
q_t^{net}	net energy bought from the grid during slot t (kWh)
p_t	time-of-use price (\$/kWh)
c_t	state of charge (SOC) i.e., amount of energy stored in BESS (kWh)
$N_b(N_s)$	number of batteries (solar panels)
$\eta_c(\eta_d)$	charging (discharging) conversion efficiency of battery (%)
η_s	efficiency of solar panel described in months
$M_t(M_p)$	number of discretization levels for electricity load (PV energy) variable
M_c	number of discretization levels for SOC of BESS
$\pi^m(i, h)$	action (charging/ discharging rate of battery) during hour h in month m at system state i
Π^m	policy matrix consisting of actions for all hours and all system states during month m

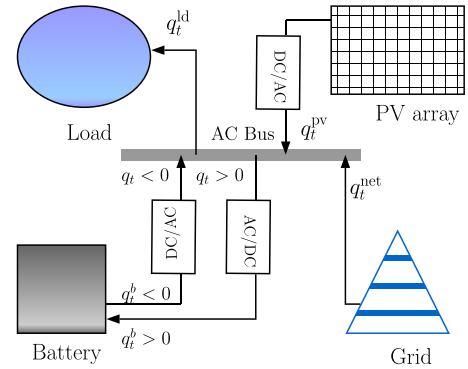


Fig. 1. Grid-connected PV system with battery and load.

2.1. Battery model

The battery model represents the dynamics of the battery regarding its mode of operation, such as idle, charging and discharging. Let c_t (in kWh) denote the state of charge (SOC) or the amount of energy stored in the BESS at the end of time slot t . Let q_t^b represent the charging/discharging power during slot t . More specifically,

- $q_t^b = 0$ if the battery is idle,
- $q_t^b > 0$ if the battery is charging, and
- $q_t^b < 0$ if the battery is discharging.

The dynamics of battery is described as

$$c_{t+1} = c_t + q_t^b. \quad (1)$$

The BESS is subject to the following two constraints:

- 1) Capacity constraint: The BESS cannot be charged above or below a certain range such that

$$N_b C'_{\min} \leq c_t \leq N_b C'_{\max} \lambda(a_b), \quad (2)$$

where N_b is the number of batteries in the BESS, $C'_{\max}$ is the initial capacity of a single battery, $C'_{\min}$ is minimum allowable capacity of a battery, and $\lambda(a_b)$ is the battery aging function defined as [17]

$$\lambda(a_b) = 1 - \alpha(a_b - 1)^{0.75} - \beta(a_b - 1)^{0.5}. \quad (3)$$

Here, α and β denote the calendar aging and cycling aging coefficients of the battery described in months, respectively, and a_b is the age of the battery in months. The aging function models the phenomenon that the maximum battery capacity gradually decreases

over a long period of time.

- 2) Charging/discharging constraint: The battery charging/discharging rate q_t^b (W) must follow

$$\max[-(c_t - C_{\min}), \bar{q}_d] \leq q_t^b \leq \min[C_{\max}(a_b) - c_t, \bar{q}_c], \quad (4)$$

where $\bar{q}_d < 0$, $-\bar{q}_d$ is the maximum battery discharging power, $\bar{q}_c > 0$ is the maximum battery charging power, $C_{\min} = N_b C'_{\min}$ is the minimum allowable capacity of BESS, and $C_{\max}(a_b) = N_b C'_{\max} \lambda(a_b)$ is the maximum capacity of BESS.

Since the battery is operated on DC, and an AC-to-DC converter is required while charging the battery, and a DC-to-AC converter is required while discharging [18] as shown in Fig. 1. We denote $\eta_c \in (0, 1]$ and $\eta_d \in (0, 1]$ as the charging and discharging conversion efficiency, respectively. For simplicity, we assume that both converters have the same conversion efficiency, i.e., $\eta_c = \eta_d = \eta$. Note that $\eta^2 = \eta_c \eta_d$ is the round trip efficiency of the battery storage. Given the battery charging/discharging rate q_t^b during time slot t , the amount of power exchanged with the AC bus q_t can be expressed as

$$q_t = \begin{cases} \eta q_t^b, & \text{if } q_t^b < 0 \\ q_t^b / \eta, & \text{otherwise} \end{cases} \quad (5)$$

2.2. PV energy and load

The PV energy and load are assumed to be exogenous stochastic processes, and we denote q_t^{pv} and q_t^{ld} as their respective quantities (in kWh) during time t . The PV energy is obtained by

$$q_t^{pv} = N_s \cdot q_t^s \cdot \eta_{pv}^{inv} \cdot \eta_s^{a_s-1}, \quad (6)$$

where N_s is the number of solar panels, q_t^s is the PV energy (DC output) from a single solar panel, η_{pv}^{inv} is the PV inverter efficiency, $\eta_s \in (0, 1]$ is the efficiency of the solar panel described in months, a_s is the age of the solar panel in months. The age information a_s can be easily calculated from the production date documented by the manufacturer of the solar panel. In our simulations, we assume both the ages of solar panel (a_s) and battery (a_b) to be zero initially, and update the ages incrementally as time goes on. The efficiency parameter η_s describes the aging effect of the solar panel over time. It is noteworthy to mention that the PV system, BESS, grid, and loads are all connected to an AC bus as shown in Fig. 1. As PV generation is operated on DC, a DC-to-AC inverter is required, and its efficiency is assumed to be a constant $\eta_{pv}^{inv} \in (0, 1]$.

It is reasonable to assume that there are different types of periodical patterns such as daily or/and weekly pattern in PV energy and electricity load profiles. So, PV energy and load demonstrate deterministic variability as well as stochastic variability. In order to model such a phenomenon, we consider the probabilistic model for periodic stochastic process [7].

Definition 1. (Periodic Stochastic Process) A stochastic process $\{X_t\}_{t=1}^{\infty}$ is periodic with period of T if the joint probability distribution of $\{X_{k+IT}\}_{k=1}^T$ is identical for all $l \in \{0, 1, \dots\}$.

For PV energy and load, the period of the periodic stochastic process is $T = 24$ hours. In addition to hours in a day, the actual distributions of the PV energy and load should also depend on the season and month in a year. Thus the probabilistic model for these features are described for each month as

$$q_t^{ld} = a_{m,h} + b_{m,h} W_t^{ld}, \quad (7)$$

$$q_t^{pv} = d_{m,h} W_t^{pv}, \quad (8)$$

where $m \in \mathcal{M} \triangleq \{1, 2, \dots, 12\}$ represents index of months in a year, $h \in \mathcal{H} \triangleq \{1, 2, \dots, 24\}$ represents index of hours in a day, $a_{m,h}$, $b_{m,h}$, $d_{m,h}$ are sets of deterministic constants defined as

$$a_{m,h} = \min_i q_{i,m,h}^{ld}, \quad (9)$$

$$b_{m,h} = \max_i q_{i,m,h}^{ld} - a_{m,h}, \quad (10)$$

$$d_{m,h} = \max_i q_{i,m,h}^{pv}. \quad (11)$$

Here, (i, m, h) corresponds to the i th observation for the h th hour in the m th month. $W_t^{ld} \in \mathcal{S}_{ld} = [0, 1]$, $W_t^{pv} \in \mathcal{S}_{pv} = [0, 1]$ are stationary and independent discrete-time Markov chains (DTMC). These parameters of the model can be estimated using historical data.

2.3. Electricity price

Electricity is purchased from the power grid at unit price r_t (\$/kWh) according to time-of-use (TOU) pricing. TOU pricing is the rate plan in which energy prices are time-dependent. TOU rates are often set in advance and kept constant throughout a contract duration. Different utilities use different time schedules for defining TOU pricing; however, they are generally classified as followings:

- Hour: peak hours, part-peak hours, and off-peak hours.
- Day: weekdays, weekends, and holidays.
- Month: summer months and winter months.

3. Problem formulation

In this section, we formulate the problem under the MDP framework.

3.1. MDP Framework

Definition 2. Markov decision process is a 4-tuple $\{S, \mathcal{A}, \mathcal{P}, \mathcal{R}\}$ in which S is a set of states, $\mathcal{A}$ is a set of actions, $\mathcal{P}$ is a transition probability function defined as $\mathcal{P}: S \times \mathcal{A} \times S \rightarrow [0, 1]$, and $\mathcal{R}$ is a reward (or cost) function defined as $\mathcal{R}: S \times \mathcal{A} \rightarrow \mathbb{R}$.

This work considers a finite MDP where the number of states and actions are finite. At each time step t , the system transitions from state $s_t \in S$ to $s_{t+1} \in S$ under the influence of a control action $a_t \in \mathcal{A}$ according to the transition probability function $\mathcal{P}$. The system incurs a cost r_t (or receives a reward) with each state transition according to $\mathcal{R}$. The goal in MDP is to find an optimum policy π^* that minimizes some form of long-term cumulative costs $\{r_t\}_{t=1}^{\infty}$. A policy π is a function that outputs an action $a \in \mathcal{A}$ for each state $s \in S$, such that $\pi: S \rightarrow \mathcal{A}$. Here, we only consider deterministic and stationary unichain policy [14]. Next, we describe the components of MDP framework according to the system model.

3.1.1. State space S

The state space S consists of two types of features.

- (i) Controllable feature: This feature contains state information related to system quantities that are influenced by the control actions. In this system model, the battery SOC $c_t \in C$ is the controllable feature. In order to formulate the problem in the form of a finite MDP with a finite number of states, the state support space C is uniformly discretized into M_c levels as

$$\Delta c = \frac{C_{\max}(a_b) - C_{\min}}{M_c - 1}, \quad (12)$$

where the step size Δc is a decreasing function of the age of BESS a_b .

- (ii) Exogenous feature: The exogenous feature includes the observable information that affects the system dynamics and the cost function, but cannot be influenced by the control actions. This feature depends on time and weather. In this case, load q_t^{ld} and PV energy q_t^{pv} are the exogenous processes that produce the features W_t^{ld} and W_t^{pv}

as defined in (7) and (8), respectively.

Note that, W_t^{ld} and W_t^{pv} are continuous since q_t^{ld} and q_t^{pv} are continuous. Therefore, in order to use the model as a finite MDP, we discretize W_t^{ld} and W_t^{pv} into M_l and M_p levels, respectively. The discretized random variables are represented as $\bar{W}_t^{\text{ld}}$ and $\bar{W}_t^{\text{pv}}$, respectively. That means, W_t^{ld} is mapped from a state space $S_{\text{ld}} = [0, 1]$ to $\bar{S}_{\text{ld}} = [0, \frac{1}{M_l-1}, \frac{2}{M_l-1}, \dots, 1]$ and W_t^{pv} is mapped from a state space $S_{\text{pv}} = [0, 1]$ to $\bar{S}_{\text{pv}} = [0, \frac{1}{M_p-1}, \frac{2}{M_p-1}, \dots, 1]$. Thus, $\bar{W}_t^{\text{ld}}$ and $\bar{W}_t^{\text{pv}}$ are the DTMCs with M_l and M_p states, respectively, and they represent the exogenous features in the state space.

To summarize, the state vector of the system at time slot t is denoted by $s_t = (c_t, \bar{W}_t^{\text{ld}}, \bar{W}_t^{\text{pv}}) \in S$, where state space $S = \{1, 2, \dots, N\}$ and the size of the state space is $N = M_c \times M_l \times M_p$.

3.1.2. Action space $\mathcal{A}$

According to the system model, the action or control variable of the MDP framework is q_t^b which denotes the charging/discharging power of the BESS during time slot t . During each time slot, the possible actions are either to leave the BESS idle or to charge the BESS or to discharge the BESS depending on the state s_t of the system. To satisfy the state transition equation of the battery SOC described in (1), the action space $\mathcal{A}$ is discretized with the same step size Δc as used in discretization of state variable c_t in (12).

$$\mathcal{A} = \{\dots, -2\Delta c, -\Delta c, 0, \Delta c, 2\Delta c, \dots\},$$

where the boundary of $\mathcal{A}$ is limited by the constraint of battery charging/discharging power in (4). Therefore, the sets of allowable actions $\mathcal{A}(s_t) \subset \mathcal{A}$ for states $s_t \in S$ can be different from each other depending on the values of state variable c_t . For a given month m , define the policy-matrix as $\Pi^m \triangleq [\pi^m(i, h)]_{i \in S, h \in \mathcal{H}}$, where $\pi^m(i, h) \in \mathcal{A}(i)$ represents the action, i.e., BESS charging/discharging power selected for state i during hour h . The dimension of the policy-matrix Π^m is $N \times 24$.

The discretization of the state and action spaces is performed to meet the requirement of a finite MDP. The discretization step size controls the tradeoff between complexity and accuracy. A finer discretization step size provides a more accurate approximation of the original system with continuous state and action spaces, but at the cost of a higher computation complexity. The impact of discretization step size on the complexity-accuracy tradeoff will be studied in Section 5 Experimental Results.

3.1.3. Transition probability $\mathcal{P}$

The transition probability $\mathcal{P}(s_t, q_t^b, s_{t+1})$ denotes the probability of the system transitioning from state s_t at time step t to state s_{t+1} at time step $t+1$ under the influence of battery charging/discharging rate q_t^b . The control variable q_t^b only affects the transition of controllable feature c_t , which is defined by the dynamics in (1). All other state variables are unaffected by q_t^b .

The stochastic variability of the system comes from the random transitions of the exogenous features represented by the DTMC variables $\bar{W}_t^{\text{ld}}$ and $\bar{W}_t^{\text{pv}}$. In order to capture the true nature of exogenous features as accurately as possible, we consider separate models of transition probability for different hours and different months. Specifically, for a given month $m \in \mathcal{M}$ and hour $h \in \mathcal{H}$, we denote $\mathcal{P}_{m,h}^{\text{ld}}$ and $\mathcal{P}_{m,h}^{\text{pv}}$ as the transition probability matrices of discretized load and PV energy, respectively. The dimensions of $\mathcal{P}_{m,h}^{\text{ld}}$ and $\mathcal{P}_{m,h}^{\text{pv}}$ are respectively $M_l \times M_l$ and $M_p \times M_p$.

For a given transition from state $s_t = (c_t, \bar{W}_t^{\text{ld}}, \bar{W}_t^{\text{pv}})$ to state $s_{t+1} = (c_j, \bar{W}_j^{\text{ld}}, \bar{W}_j^{\text{pv}})$ under action $q_t^b = a$, we denote $p_{ij}^{m,a}(h)$ as the transition probability for the system in month m and in hour h . According to the system model,

$$p_{ij}^{m,a}(h) = \begin{cases} 0, & \text{if } c_j \neq c_t + a \\ \mathcal{P}_{m,h}^{\text{ld}}(\bar{W}_t^{\text{ld}}, \bar{W}_j^{\text{ld}}) * \mathcal{P}_{m,h}^{\text{pv}}(\bar{W}_t^{\text{pv}}, \bar{W}_j^{\text{pv}}), & \text{o.w.} \end{cases}$$

We define $\mathcal{P}^m \triangleq [p_{ij}^{m,a}(h)]_{i,j \in S, a \in \mathcal{A}(i), h \in \mathcal{H}}$ as the transition probabilities of the overall system in month m .

For a given policy-matrix $\Pi^m = [\pi^m(i, h)]_{i \in S, h \in \mathcal{H}}$, let denote $\mathbf{P}_H^m(h) = [p_{ij}^{m,\pi^m(i,h)}(h)]_{i,j \in S}$ as the transition probability matrix at hour h of month m . $\mathbf{P}_H^m(h)$ is a stochastic matrix with the dimension $N \times N$.

Therefore, the intermittent variations in load and PV energy can be effectively captured by using the transition probabilities of the periodic MDP framework. The optimum solution to the MDP problem depends on these transition probabilities, which model the stochastic and intermittent nature of load and PV energy.

3.1.4. Cost function $\mathcal{R}$

The system cost function $\mathcal{R}$ is used to evaluate the system performance under a given policy π . In our system model, the price of the electricity is used as the performance metric. During each time slot t , for a given load q_t^{ld} , PV energy q_t^{pv} , and charging/discharging rate q_t^b , the net energy bought from the grid can be expressed as

$$q_t^{\text{net}} = q_t^{\text{ld}} - q_t^{\text{pv}} + q_t^b, \quad (13)$$

where q_t is defined as a function of q_t^b in (5). So, the cost of electricity purchased from the grid in order to meet demand at each time slot t (corresponding to month m and hour h) can be computed as

$$\mathcal{R}_t^m(s_t, q_t^b) = \max(0, q_t^{\text{net}}) \cdot r_t^m, \quad (14)$$

where r_t^m is the per-unit TOU electricity price (with slight abuse of notation) in month m and hour $h = \text{mod}(t-1, 24) + 1$, with $\text{mod}(\cdot)$ being the modulo operator. If the system originates from state i at hour h in month m , the expected immediate cost with action a can be expressed as

$$g_i^{m,a}(h) = \sum_{j \in S} p_{ij}^{m,a}(h) \mathcal{R}_h^m(i, a) = \mathcal{R}_h^m(i, a). \quad (15)$$

3.2. Optimization objective

We consider the long-term average cost of electricity as the objective function. For a given stationary unichain policy Π^m , the average cost of electricity that is drawn from the grid in month m can be defined as

$$R(\Pi^m) \triangleq \mathbb{E} \left[\sum_{t=1}^{\infty} \gamma^{t-1} \mathcal{R}_t^m(s_t, q_t^b) \right], \quad (16)$$

where $\gamma \in (0, 1)$ is a discount factor, and $\mathbb{E}[\cdot]$ represents expectation taken with respect to the measure induced by the random DTMC processes $\bar{W}_t^{\text{ld}}$, $\bar{W}_t^{\text{pv}}$, and the policy Π^m .

Based on the above definitions, the dynamic battery scheduling problem can be formulated as

$$(P1) \quad R_m^* \triangleq \min_{\Pi^m} R(\Pi^m)$$

where R_m^* denotes the minimum long-term discounted accumulative electricity cost obtained from the optimum policy in month m . According to [19], (P1) is a nonstationary discrete-time periodic MDP. The objective is to find the optimum policy-matrix Π^{m*} for every month $m \in \mathcal{M}$ separately, such that the average discounted electricity purchase cost is minimized.

4. Periodic MDP-based online battery scheduling

In this section, we propose to solve the periodic MDP defined in (P1) by using dynamic programming. The results are then used for online battery management in real-world systems.

For any periodic MDP, there exist value functions $v_i^{m*}(h)$, $\forall i \in S$, $h \in \mathcal{H}$ and a scalar $\bar{R}_m^*$ satisfying the Bellman optimality equation [20]

Input: State space $\mathcal{S}$, action space $\mathcal{A}(i)$ for all states $i \in \mathcal{S}$, transition probabilities $\mathcal{P}^m$, cost function $\mathcal{R}(s, a) \forall s \in \mathcal{S}, a \in \mathcal{A}_s$, index of month $m \in \mathcal{M}$.

1: **Initialization:** Initialize the iteration step $k = 0$ and initialize the policy-matrix Π_0^m with actions that minimize the immediate cost.

2: **do**

3: Set $k \leftarrow k + 1$ and $h = 1$.

4: Calculate $w_i^m(h)$ by using (19) and Π_{k-1}^m .

5: Solve the following set of N linear equations in N variables $\{v_i^m(h)\}_{i \in \mathcal{S}}$

$$\left\{ v_i^m(h) = w_i^m(h) + \gamma^{24} \sum_{j \in \mathcal{S}} p_{ij}^m(h; 24) v_j^m(h) \right\} \forall i \in \mathcal{S}.$$

6: **for** h' from 24 to 1 **do**

7: Update the value functions and policy-matrix.

$$v_i^m(h') = \min_{a \in \mathcal{A}(i)} \left[g_i^{m,a}(h') + \gamma \sum_{j \in \mathcal{S}} p_{ij}^{m,a}(h') v_j^m(h' \oplus 1) \right],$$

$$\Pi_k^m(i, h') = \operatorname{argmin}_{a \in \mathcal{A}(i)} \left[g_i^{m,a}(h') + \gamma \sum_{j \in \mathcal{S}} p_{ij}^{m,a}(h') v_j^m(h' \oplus 1) \right],$$

8: **end for**

9: **while** $\Pi_k^m \neq \Pi_{k-1}^m$

Output: Policy-matrix Π_k^m for m th month.

Algorithm 1. Policy iteration algorithm for periodic MDP.

$$v_i^{m*}(h') + R_m^* = \min_{a \in \mathcal{A}(i)} \left[g_i^{m,a}(h') + \gamma \sum_{j \in S} p_{ij}^{m,a}(h') v_j^{m*}(h) \right] \quad (17)$$

such that the policy-matrix Π^{m*} resulting from the optimum value functions achieves the minimum cost R_m^* for month m . Here, $h' = 24$ if $h = 1$, and $h' = h - 1$ otherwise.

From the Bellman equation in (17), it can be observed that Π^{m*} depends on the state $i \in S$ through the value function $v_i^{m*}(h)$. Obtaining the value functions involves solving the Bellman equations for all states $i \in S$ and hours $h \in \mathcal{H}$, for which there is no closed-form solution in general [20]. Exact solution to MDP can be obtained either by linear programming (LP) based methods or dynamic programming (DP) based methods such as value iteration (VI) [16,21] or policy iteration (PI) [22]. DP-based methods utilize iterative approach that is more efficient than LP-based solution in case of large-scale discrete-time MDPs.

The value iteration method does not converge for systems with periodic transitions [14, Ch. 8.5.1]. In order to overcome this issue, the transition probabilities $\mathcal{P}$ and the cost functions $\mathcal{R}$ can be transformed in a modified VI algorithm to achieve an equivalent aperiodic MDP [14, Ch. 8.5.4]. However, the modified value iteration algorithm suffers from low convergence speed in case of high dimensional state and action spaces. Therefore, in this work, we adopt the policy iteration algorithm for periodic MDP [19] to solve the Bellman equation and obtain the policy-matrix Π^m for every month $m \in \mathcal{M}$.

4.1. Policy iteration algorithm for periodic MDP

The conventional PI procedure [22] converges to an optimum policy in the case of stationary finite Markov processes, by iteratively selecting a better policy based upon the results from the former policy in the previous iteration. A similar method is developed in Riis [19] for non-stationary periodic Markov processes. In order to examine the effect of each stage in a periodic MDP separately, the periodic PI procedure deals with a policy or decision-matrix that is divided into T stages, with T denoting the period in a periodic MDP ($T = 24$ in our case).

Based on the 24-hour non-stationary periodic feature of the problem, we introduce the following definitions to facilitate the implementation of the PI method. First, define the l -step transition probability matrix as

$$\mathbf{P}_H^m(h; l) \triangleq \prod_{k=0}^{l-1} \mathbf{P}_H^m(h+k), \quad (18)$$

where $\mathbf{P}_H^m(h) = [p_{ij}^{m,\pi^m(i,h)}(h)]_{i,j \in S}$ is the one-step transition probability matrix at hour h of month m $\mathbf{P}_H^m(h+k)$ under policy-matrix $\Pi^m = [\pi^m(i, h)]_{i \in S, h \in \mathcal{H}}$, $\mathbf{P}_H^m(h; l) \triangleq [p_{ij}^m(h; l)]_{i,j \in S}$ is a stochastic matrix, and $p_{ij}^m(h; l)$ denotes the probability of the system transitioning from state i at hour h to state j at hour $h+l$ in month m for a given policy-matrix Π^m .

For a given policy-matrix Π^m , denote $\mathbf{g}_H^m(h) = [g_i^{m,\pi^m(i,h)}(h)]_{i \in S} \in \mathbb{R}^{N \times 1}$ as the expected immediate cost vector at hour h of month m . Then, for a system starting from hour h at month m , the expected total cost over a 24-hour period can be defined as

$$\mathbf{w}_H^m(h) = \mathbf{g}_H^m(h) + \sum_{k=1}^{23} \gamma^k \mathbf{P}_H^m(h; k) \mathbf{g}_H^m(h \oplus k), \quad (19)$$

where $h \oplus k = \text{mod}(h+k-1, 24) + 1$, $\mathbf{w}_H^m(h) \triangleq [w_i^m(h)]_{i \in S} \in \mathbb{R}^{N \times 1}$ and $w_i^m(h)$ denotes the expected total cost over a 24-hour period given that the system started in state i at hour h at month m under policy-matrix Π^m .

With the above definitions, for a given policy matrix Π^m , the value function of a non-stationary periodic Markov process with $T = 24$ can be expressed as

$$v_i^m(h) = w_i^m(h) + \gamma^{24} \sum_{j \in S} p_{ij}^m(h; 24) v_j^m(h), \quad \forall i \in S. \quad (20)$$

In (20), the value of $w_i^m(h)$ can be calculated by using (19) and the policy-matrix Π^m . Given fixed Π^m , m , and h while changing i , the above equation results in a system of N linear equations with N unknown variables $\{v_i^m(h)\}_{i \in S}$, where $N = |S| = M_c \times M_l \times M_p$ is the size of state space. Thus for a given m and h , we can initialize the value of $v_i^m(h)$ at the end of a 24-hour period by solving the equation system defined by (20). The initialized value function can then be applied to the Bellman's Eq. (17) to update the policy in a 24-hour period. Details of the algorithm are given in Algorithm 1.

It is proven in Riis [19] that the value functions $v_i^{m*}(h)$ is improved for at least one state in every iteration, and the PI procedure converges to an optimum policy-matrix for a finite periodic MDP.

The value function in (20) is developed for a non-stationary periodic Markov process with a period $T = 24$. Thus the time horizon for optimization is 24 h. The output of the proposed Algorithm 1 is a policy matrix Π^m of $N \times 24$ dimension, which provides the battery charging/discharging rates in different states and different hours of a day in the m th month. As each day progresses, the system model parameters and transition probabilities of the MDP get updated according to the newly observed load and solar energy data. Details of the parameter update process will be discussed in the next subsection.

4.2. Online implementation

Based on the PI procedure listed in Algorithm 1, the online implementation of the optimum battery scheduling is given in Algorithm 2. In practical scenarios, the transition probability function $\mathcal{P}^m$ evolves with respect to time. Thus the scheduling algorithm should be dynamically adjusted based on the update of the transition probability function. The PI procedure given in Algorithm 1 runs based on a fixed probabilistic model defined by transition probability function $\mathcal{P}^m$. The algorithm provides policy-matrix Π^m , i.e., the battery charging/discharging decisions as a function of state variables for a 24-hour period in a particular month m . In order to implement an online (real-time) battery charge scheduling scheme, the system model needs to be daily re-calibrated by updating the transition probability function $\mathcal{P}^m$ based on newly collected data of load and PV energy, and then solve it for operations on the next day in a rolling horizon fashion. Similar approach is also adopted in Hu and Defourny [13]. This update will help to model the stochastic behavior of the system more accurately, and thus yield better policy. Details of the online implementation of the optimum battery scheduling are summarized in Algorithm 2.

5. Experimental results

In this section, experimental results are presented to demonstrate the performance of the proposed MDP-based battery scheduling algorithm. The results are based on the load data from a large hotel in San Francisco (SF) [23]. The PV energy is obtained by using the PVWatts calculator from National Renewable Energy Laboratory (NREL) [24] with the weather information of San Francisco. The solar system size is limited by the area of the hotel rooftop, and the number of 10 kW solar panels is set at $N_s = 120$. The price of each solar panel is set at \$6400, including the price of products and installation. The PV inverter efficiency is $\eta_{pv}^{inv} = 96\%$ and the storage efficiency for solar panels described in month is $\eta_s = 99.96\%$ [25]. The electricity charges are calculated by using the TOU rate (p_t) of PG&E [26], which is given in Table 2 along with the time schedule.

Tesla Powerwall batteries [27] are used to model the BESS, which have a 5 kW charging/discharging rate ($\bar{q}_c/\bar{q}_d$) and 13.5 kWh capacity (C'_{max}) each. Price of each battery is \$5900 and the conversion efficiency of the batteries is $\eta = 94\%$. According to the datasheet [27], the calendar aging coefficient is $\alpha = 0.0036$, and the cycling aging coefficient

Input: Discretization levels for load M_l , for PV energy M_p , and for BESS SOC M_c .

1: **Initialization:** Denote initial time index as $t = 0$, initialize the parameters for the probabilistic model, i.e., $a_{m,h}$, $b_{m,h}$, $d_{m,h}$, and the transition probability functions $\mathcal{P}^{\text{old}}$ and $\mathcal{P}^{\text{pv}}$ based on the training data.

2: **while** True **do**

3: Given time index t (in hours), determine month index m , age of battery a_b , and age of solar panels a_s .

4: Solve the optimum policy-matrix Π^m by using Algorithm 1 with the current transition probability function.

5: **for** $k = 1$ to 24 **do**

6: Collect newly observed test data q_{t+k}^{ld} and q_{t+k}^{pv} , and observe the SOC of BESS c_{t+k} .

7: Given time index $t + k$, determine the hour index h and month index m .

8: From state space S , identify the state i according to equations (7) to (11) and the values of c_{t+k} , M_l , M_p , M_c .

9: Apply the battery charging/discharging power $q_{t+k}^b = \Pi^m(i, h)$

10: **end for**

11: According to $\{q_{t+k}^{\text{ld}}\}_{k=1}^{24}$ and $\{q_{t+k}^{\text{pv}}\}_{k=1}^{24}$, update the model parameters and transition probabilities.

12: Set $t \leftarrow t + 24$.

13: **end while**

14: Return to Step 2.

Algorithm 2. Proposed online battery charge scheduling.

Table 2

Time schedule and TOU rates (\$ per kWh) in San Francisco [26].

	Summer (5/1 to 10/31)			Winter (11/1 to 4/30)	
	Peak	Part-peak	Off-peak	Part-peak	Off-peak
Time	12p-6p	9a-12p, 6p-9p	9p-9a	9a-9p	9p-9a
TOU	0.15384	0.11333	0.08651	0.10779	0.09317

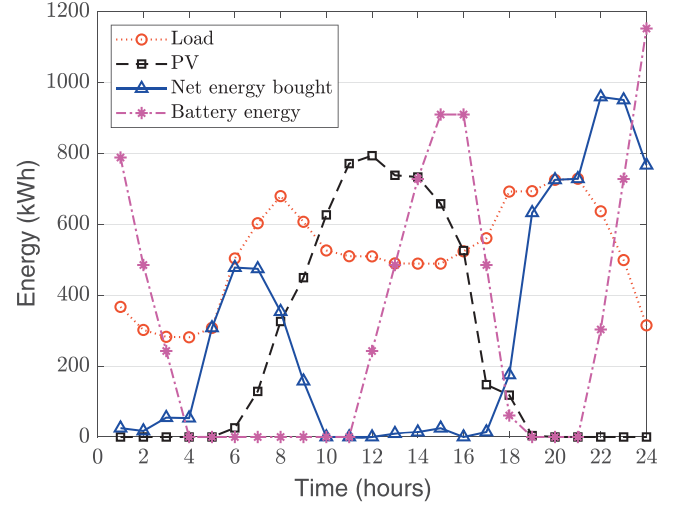


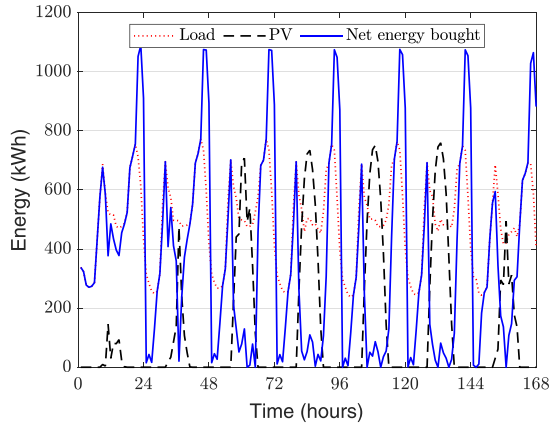
Fig. 2. Snap shot of 1-day energy usage on July 1st.

is $\beta = 0.0155$. The initial SOC of BESS is 0. The number of batteries N_b are set at 90. The discretization levels for load, PV energy, and SOC of BESS are set as $M_l = 6$, $M_p = 6$, and $M_c = 10$, respectively (unless otherwise specified). The discount factor is $\gamma = 0.999$.

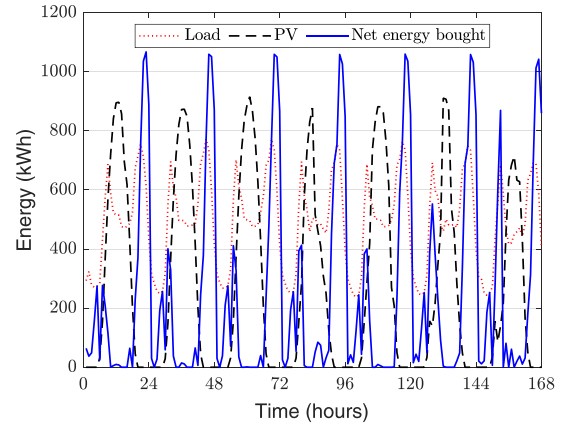
Fig. 2 illustrates a snapshot of 1-day energy usage on July 1st. The energy profiles are shown as functions of the hour of the day. The proposed solution yields the values of the net energy bought from grid q_t^{net} and battery SOC c_t . During 10:00 to 16:00, the PV energy exceeds the load and the amount of energy bought is almost zero. The extra energy from PV is used to charge the BESS during this period. During late night 21:00 to 24:00, extra energy is bought from the grid even after meeting the load demand in order to utilize the low TOU rate during those off-peak hours. As a result, these extra energy is used to charge the BESS and later used to meet the load demand during peak hours, when TOU rate is much higher. Therefore, considerable amount of utility cost is saved with the proposed periodic MDP-based solution.

The snap shots of 1-week energy usage in the first week of a typical winter month and a typical summer month are shown in Fig. 3. The loads in both months are similar due to the relatively mild weather in San Francisco. There are usually two peaks in the load profile in every day of the week: the early one is around 8:00 and the later one is around 20:00. Also, the peak in PV energy profile during mid-day is higher in summer than in winter. The distinguishing effect of the proposed algorithm in these two weekly energy profiles can be observed in the amount of net energy bought from the grid during early morning. In January, the net energy exceeds 600 kWh in most of the days, while the net energy bought in June is under 400 kWh mostly. Even though the load profile is similar in both months, due to higher PV energy peak value in June, the demand during early morning hours can be met by utilizing the energy stored in BESS.

Table 3 compares the result of the proposed online battery charge scheduling scheme with the optimum results obtained from offline optimization [6], the results obtained from the One-step Roll-out algorithm (ORA) proposed in Kwon et al. [7], and the results obtained from the rule-based scheme (RS) presented in Bhende et al. [8]. In the offline optimization scheme, the system has complete knowledge of



(a) First week of January



(b) First week of June

Fig. 3. Snap shots of 1-week energy usage in the first weeks of January and June, respectively.

future data on load and PV power over an entire year, yet these data are not available in practical systems. Thus the results of the non-causal offline scheduling algorithm serve as a non-achievable lower bound for any practical algorithms. The ORA is a heuristic policy based on approximate dynamic programming, which approximates the cost-to-go by solving a deterministic (certainty-equivalent) optimization problem with all random variables taking their respective expected values. The RS algorithm provides dynamic scheduling by using decision trees with the help of forecast load and PV energy. The proposed online scheduling algorithm, ORA, and RS only have knowledge of past data. For ORA and the proposed algorithm, the numbers of all discretization levels are set as $M_l = 6$, $M_p = 6$, and $M_c = 10$, while the data from the previous day is used as the forecast data in RS. In spite of not having access to the actual data from the future, the performance of the proposed online scheduling algorithm is close to that of the non-causal offline algorithm. Compared to a system without PV or BESS, the non-causal offline algorithm achieves a 43.86% saving in the annual bill. In comparison, our proposed online solution yields a 41.68% saving in the annual bill. The annual bill from the proposed online algorithm is only 3.88% higher than that from the non-causal offline algorithm. The savings achieved by the ORA and RS algorithms are 39.80% and 40.64%, respectively.

Fig. 4 demonstrates the impacts of discretization levels on the relative difference in annual electricity bill with respect to the non-causal offline lower bound of \$264,510. Intuitively, as the numbers of all discretization levels M_b , M_p , and M_c increase, the result of the proposed solution gets better, since high-dimensional state space provides a better representation of the system dynamics. On the other hand, the computational complexity of the proposed algorithm increases with the number of discretization levels, which raises the issue of tradeoff between complexity and performance of the algorithm. According to the results in Fig. 4, the difference in the performance of the proposed algorithm is very small between $M_c = 10$ and $M_c = 20$. Therefore, after a certain point, increasing the values of discretization level will not yield any significant improvement in performance.

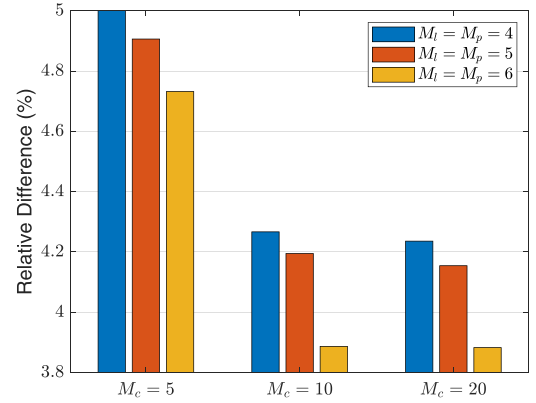


Fig. 4. Effect of discretization levels on relative difference with respect to the non-causal offline lower bound of annual electricity bill of \$264,510.

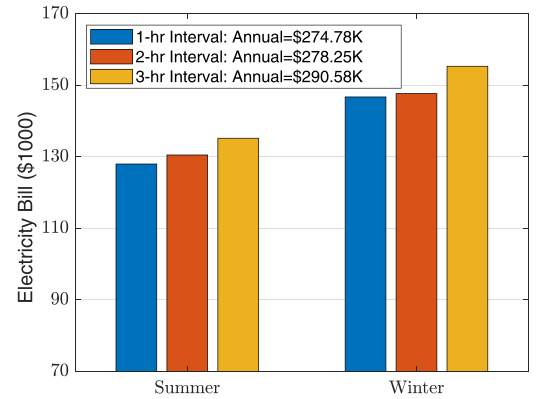


Fig. 5. Effects of time discretization interval on electricity bill.

Table 3

Annual electricity bill under BESS-assisted PV system.

System	Proposed online	Optimum offline [6]	ORA [7]	RS [8]
Electricity Bill (\$)	274,780	264,510	283,620	279,690
Bill without BESS & PV (\$)	471,160			
Savings (\$)	196,380 (41.68%)	206,650 (43.86%)	187,540 (39.80%)	191,470 (40.64%)
System Cost (\$)	1,299,000			
Break-even point (months)	80	76	84	82

Table 4

Average computation times of the proposed algorithm (in seconds) under different system configurations.

M_c	1-hour interval	2-hour interval	3-hour interval
5	1.28	0.53	0.38
10	4.26	2.05	1.42
20	32.34	23.05	14.89

Fig. 5 illustrates the effects of time discretization interval on the performance of the proposed solution. In all previous results, the time is discretized into 1-hour intervals. The complexity of the algorithm can be further reduced by increasing the discretization time interval. In the results in Fig. 5, the annual electricity bill is divided into two separate periods of the year, summer and winter (c.f. Table 2). Intuitively, a longer time interval will result in a lower time resolution during optimization, which in turn leads to a decrease in performance. Specifically, the system model with 1-hr interval achieves 1.25% and 5.44% better performance than the models with 2-hr and 3-hr intervals, respectively.

Table 4 shows the average computation times of the proposed algorithm (Algorithm 2) under different configurations of M_c and time discretization interval. The results are obtained on a computer with an Intel i5 dual-core 2.2 GHz processor with 6 GB RAM and calculated by taking the daily average. It is worth noting that the majority of computation need is required for solving the optimum policy matrix, which is performed only once every day. The computation time is on the order of dozens of seconds or less. The results justify that the proposed battery charge scheduling algorithm can be implemented in real-time with negligible delay. It can be observed that the algorithm requires more computation time with the increase in discretization level M_c due to the larger number of states. Moreover, as the time discretization interval increases, the algorithm requires less data to process, thus yielding smaller computation time.

6. Conclusion

The optimum online BESS management policy has been studied in this paper for grid-connected PV systems. The policy was designed to identify the optimum charging/discharging schedule of BESS so that the long-term cost of energy purchased from the grid is minimized. The optimization problem was formulated by using the framework of periodic discounted MDP and solved by using a policy iteration based approach. The aging effects of the batteries and solar panels, and the stochastic variability in loads and PV energy have been included in the problem formulation. Simulation results demonstrated that the proposed algorithm can achieve a 41.68% saving in annual utility bills compared to conventional systems without BESS or PV. In addition, the annual bill resulted from the proposed online algorithm is only 3.88% higher than the optimum lower bound obtained from a non-causal offline algorithm.

For future works, we can further improve the performance of the system by employing model-free approaches such as deep reinforcement learning (DRL), which does not require an explicit model of the transition probabilities. In addition, one particular DRL approach, deep deterministic policy gradient (DDPG), can deal with continuous state space, thus further improve the precision of the results.

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.

References

- [1] P. Harsha, M. Dahleh, Optimal management and sizing of energy storage under dynamic pricing for the efficient integration of renewable energy, *IEEE Trans. Power Syst.* 30 (3) (2015) 1164–1181.
- [2] Y. Riffonneau, S. Bacha, F. Barruel, S. Ploix, et al., Optimal power flow management for grid connected PV systems with batteries, *IEEE Trans. Sustain. Energy* 2 (3) (2011) 309–320.
- [3] J. Teng, S. Luan, D. Lee, Y. Huang, Optimal charging/discharging scheduling of battery storage systems for distribution systems interconnected with sizeable PV generation systems, *IEEE Trans. Power Syst.* 28 (2) (2013) 1425–1433.
- [4] S. Zhang, Y. Tang, Optimal schedule of grid-connected residential PV generation systems with battery storages under time-of-use and step tariffs, *J. Energy Storage* 23 (2019) 175–182.
- [5] L. Luo, S.S. Abdulkareem, A. Rezvani, M.R. Miveh, S. Samad, N. Aljojo, M. Pazhoohesh, Optimal scheduling of a renewable based microgrid considering photovoltaic system and battery energy storage under uncertainty, *J. Energy Storage* 28 (2020) 101306.
- [6] Y. Li, J. Wu, Optimum design of battery-assisted photo-voltaic energy system for a commercial application, *Proc. IEEE Power & Energy Society General Meeting (PESGM)*, Atlanta, GA, USA, (2019).
- [7] S. Kwon, Y. Xu, N. Gautam, Meeting inelastic demand in systems with storage and renewable sources, *IEEE Trans. Smart Grid* 8 (4) (2017) 1619–1629.
- [8] C.N. Bhende, S. Panda, S. Mishra, A. Narayanan, T. Kaipia, J. Partanen, Optimal power flow management and control of grid connected photovoltaic-battery system, *Int. J. Emerg. Electr. Power Syst.* 20 (5) (2019) 20190056.
- [9] A. Rezvani, M. Gandomkar, M. Izadbakhsh, A. Ahmadi, Environmental/economic scheduling of a micro-grid with renewable energy resources, *J. Clean. Prod.* 87 (2015) 216–226.
- [10] Y. Li, S.Q. Mohammed, G.S. Nariman, N. Aljojo, A. Rezvani, S. Dadfar, Energy management of microgrid considering renewable energy sources and electric vehicles using the backtracking search optimization algorithm, *J. Energy Resour. Technol.* 142 (5) (2020) 052103.
- [11] C. Liu, S.S. Abdulkareem, A. Rezvani, S. Samad, N. Aljojo, L.K. Foong, K. Nishihara, Stochastic scheduling of a renewable-based microgrid in the presence of electric vehicles using modified harmony search algorithm with control policies, *Sustain. Cities Soc.* 59 (2020) 102183.
- [12] B. Mbuwir, F. Ruelens, F. Spiessens, G. Deconinck, Battery energy management in a microgrid using batch reinforcement learning, *Energies* 10 (11) (2017) 1846.
- [13] Y. Hu, B. Defourny, Near-optimality bounds for greedy periodic policies with application to grid-level storage, *Proc. IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, Orlando, FL, USA, (2014).
- [14] M.L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, 2014.
- [15] D.K. Maly, K.-S. Kwan, Optimal battery energy storage system (bess) charge scheduling with dynamic programming, *IEE Proc. Sci. Meas. Technol.* 142 (6) (1995) 453–458.
- [16] S. Nath, J. Wu, H. Lin, Optimum multicast scheduling in delay-constrained content-centric wireless networks, *Proc. IEEE International Conference on Communications (ICC)*, Shanghai, China, (2019).
- [17] A. Berrueta, J. Pascual, I. San Martín, P. Sanchis, A. Ursúa, Influence of the aging model of lithium-ion batteries on the management of PV self-consumption systems, *IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe (IEEEIC/IE&CPS Europe)*, (2018).
- [18] Y. Ru, J. Kleissl, S. Martinez, Storage size determination for grid-connected photovoltaic systems, *IEEE Trans. Sustain. Energy* 4 (1) (2013) 68–81.
- [19] J.O. Riis, Discounted Markov programming in a periodic process, *Oper. Res.* 13 (6) (1965) 920–929.
- [20] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, 2 Athena scientific Belmont, MA, 2011.
- [21] D.J. White, Dynamic programming, Markov chains, and the method of successive approximations, *J. Math. Anal. Appl.* 6 (3) (1963) 373–376.
- [22] M.L. Littman, T.L. Dean, L.P. Kaelbling, On the complexity of solving Markov decision problems, *Proc. 11th conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1995, pp. 394–402.
- [23] Commercial and residential hourly load profiles for all TMY3 locations in the united states, <https://openai.org/datasets/dataset>.
- [24] Pvwatts calculator, <https://pvwatts.nrel.gov/pvwatts.php>.
- [25] C.A. Fernandes, J.P.N. Torres, M. Morgado, J.A. Morgado, Aging of solar PV plants and mitigation of their consequences, *Proc. International Power Electronics and Motion Control Conference (PEMC)*, IEEE, 2016, pp. 1240–1247.
- [26] Electric schedule e-19-medium general demand - metered TOU service, https://www.pge.com/tariffs/tm2/pdf/ELEC_SCHEDULE_E-19.pdf.
- [27] Tesla powerwall battery, <https://www.tesla.com/powerwall>.