Optimal Storage Scheduling Using Markov Decision Processes

Samuele Grillo, *Member, IEEE*, Antonio Pievatolo, and Enrico Tironi

***Abstract*—The paper presents a method based on Markov decision processes to optimally schedule energy storage devices in power distribution networks with renewable generation. The time series of renewable generation is modeled as a Markov chain which allows for the implementation of a stochastic dynamic programming algorithm. The output of this algorithm is an optimal scheduling policy for the storage device achieving the minimization of an objective function including cost of energy and network losses. Besides this, other properties, such as energy storage placement and size, can be assessed and compared in optimized systems with different layouts.**

***Index Terms*—Energy storage, Markov processes, renewable generation, stochastic dynamic programming.**

Nomenclature

t	index of an epoch
γ_t	load
π	a control policy
a_t	energy drawn from, or supplied to, the storage device
b_t	state of charge
c_t	clearness
d_t	decision function from the state space to the action space
E	nominal capacity of the storage device
n_{clr}	number of clearness levels
n_{SOC}	number of states of charge
r_t	reward function
s_t	system state
$u_t^{\pi}(s_t)$	value-to-go at epoch t for policy π and current system state s_t
w_t	system energy losses
z_t	price of energy

I. Introduction

THE increasing presence of electrical storage systems in power networks requires to optimize their management [1], [2]. Network losses and the price of energy are two key factors that must be taken into account when solving this decision problem at a minimum cost, with the additional complication of randomness of both the power supply (especially if provided by renewable sources) and the load.

There are several contributions to the coordinated management of energy storage devices and renewable sources. [3]–[6] consider the renewable generation profile as known with no uncertainty. As for the algorithmic approach to storage management optimization, [3], [4] use dynamic programming, whereas in [5], [6] the optimization is performed simultaneously for all decision epochs by means of mixed integer linear programming and particle swarm optimization, respectively. However, in order to include the intermittent nature of renewable energy sources and the sequential character of the decision problem in the analysis, a stochastic dynamic programming approach is to be preferred. Therefore, in this work, the optimal storage management policy (the optimal policy from now on) is based on Markov decision processes (MDP, see [7]), assuming a random renewable energy source connected to the grid and a deterministic load profile. We solved a 24-hour finite-horizon optimal policy problem, in which the solar energy produced every 15-minute interval is obtained through a Markov chain model of the clearness, as done by [8].

The Markov chain framework is attractive for storage management policy because the state of charge (SOC) at discrete time steps can be regarded as a Markov chain. In fact, the SOC at any time, given all the past SOCs, can be written as a function of the previous SOC and some random variables which do not depend on the past. The surrounding processes, which cause the SOC to change, can be non-Markov, however a Markov chain could approximate the true process satisfactorily enough. This viewpoint is shared by other authors. For example in [9] an MDP approach to the management of the batteries of a solar-powered sensor network which maximizes the ability of the sensor node to detect and transmit an event of interest, while aiming at preserving the battery energy level, has been proposed. In [10] an enhanced stochastic dynamic programming for the minimization of electric vehicles' charging costs in a smart grid framework. The enhancement consists in the possibility of using a continuous space of decision variables. In [11] authors addressed the problem of organizing energy storage purchases (for households and data centres) to minimize long-term energy costs under variable demands and prices in an MDP framework. They found that the optimal policy has a two-threshold structure, that is, the battery must be discharged when the SOC is above a certain threshold and must be charged when the SOC is below another threshold. They have also briefly suggested how to incorporate a renewable energy source and

Manuscript received April 15, 2015; revised August 09, 2015 and October 07, 2015; accepted October 30, 2015. Date of publication November 26, 2015; date of current version March 18, 2016. Paper no. TSTE-00294-2015.

S. Grillo and E. Tironi are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, I-20133 Milan, Italy (e-mail: samuele.grillo@polimi.it; enrico.tironi@polimi.it).

A. Pievatolo is with CNR-IMATI, I-20133 Milan, Italy (e-mail: antonio.pievatolo@mi.imati.cnr.sit).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

battery replacement costs within the optimization problem, but have not implemented the suggested solution. In [12] a simplified solution to the same problem has been obtained, by considering the price the only (Markovian) stochastic quantity and by relaxing ramping constraints on storage charging and discharging; they found an optimal policy based on a threshold for the price of energy. The distinctive feature of the MDP methodology is that it guarantees that no policy can obtain a better performance under non-restrictive assumptions, as shown for example by [7, Section IV.4]. Coming to the subject matter of this work, [13] have proven that a dynamic offset policy for storage scheduling based on a MDP outperforms heuristic methods, such as those aiming at maintaining a constant level of the storage.

In [14], the authors addressed the day-ahead wind commitment problem with storage, in a stochastic dynamic programming framework. At each decision epoch, the wind farm manager must pledge the electricity that the wind farm must produce the next 24 hours, knowing the 24 hourly market prices of electricity. It is assumed that the transition of the state variable, which includes the wind as its only stochastic part, is Markovian, reducing their problem to the MDP framework, with a continuous state space and in high dimension. The proposed solution is based on a discretization of the state-space obtained from a finite sample of sequences of states over the decision horizon and on a convex approximation of the value-to-go function, so that it can be easily optimized for each state of the system.

Our approach belongs to the same framework of [14], however it solves the storage scheduling optimization problem exactly, because the state and action spaces are both discrete and low dimensional: after executing the policy optimization, one is left with a decision table which, for every decision epoch, matches each state of the system (made up of both the actual SOC of the storage device and the observed radiation) with an optimal action, i.e., how much energy should be exchanged between the grid and the storage device. The optimal policy is associated with an expected revenue. Other differences can be found in the emphasis given to the various facets of the problem. The renewable source is the solar energy, which, more than wind, is correlated with the load, because most of the energy demand occurs during daylight hours. The effect of this relationship will appear in the optimal storage scheduling solution, in which the charge and discharge cycles tend to be anticyclic with respect to the price of energy, in the test case we have examined. Finally, we have also focused on the number of full storage cycles, having found that the optimization with respect to cost of energy and network losses has the side effect of reducing this number, in certain system configurations [15], while the differences between the optimal cost and the cost associated with other solutions are not large, in absolute terms. Other authors have also considered network losses as a performance measure. For example, deterministic dynamic programming has been used by [16] to define an optimal control strategy of distributed energy storages controlled by a distribution system operator. The optimality criterion is to minimize network losses, which are then used to compare different distributed energy storage placements and ratings.

The method illustrated in this article can be employed in at least two ways. The first way as an analysis tool for use by a utility company, for example for assessing what one can gain by including a storage in the system, by comparing the optimal revenue with the storage against that without it. In this case, the Markov model of the clearness for representative periods in the year is estimated from historical data. In the second way, this method provides an actual decision tool for one-day-ahead storage system management planning, where the Markov model for the clearness would rather be obtained from weather forecasts for the next day. Here we have analysed the first usage, evaluating the just mentioned optimal performance measures (revenues and number of cycles) for different storage system sizes.

The plan of the paper is as follows. In Section II we describe the MDP framework applied to our problem, we define our objective function and the action space, and we illustrate the dynamic programming algorithm, which has been obtained by joining a deterministic (during night hours) and a stochastic part (during solar production hours). In Section III we explain our case study, introducing the test network and the parameter used for the solar radiation model based on the clearness. In Section IV we describe our results. Section V contains concluding remarks, emphasising uses to which the algorithm could be put to, other than simple storage scheduling.

II. The Setup for Optimal Policy Finding

In this section we illustrate the main variables and the method for finding an optimal policy.

As done by [8], the solar power at epoch t is expressed in terms of the clearness c_t, which is the square root of the ratio between the observed radiation and the expected radiation. The advantage of this approach is that daily and seasonal patterns in both the mean and the variability of solar radiation are removed and the clearness can be modelled as a stationary autoregressive Gaussian time series, with a bimodal marginal distribution, with one primary mode representing clear sky and a less peaked secondary mode representing a range of cloud cover conditions. From Glasbey's model [8] we derived a Markov chain discretizing the clearness to 14 levels, because conditions for the existence of an optimal Markovian policy are more easily met when the state space is countable (see [7, Proposition 4.4.3]). The Markov model holds from 8 am to 4 pm. During the remaining hours no production of solar energy is assumed.

In the MDP framework one is required to specify a state vector, an action space and a reward function. We denote the state at epoch t as $s_t = (c_t, b_t, \gamma_t, z_t)$, where c_t is the clearness, b_t is the SOC, γ_t is the load and z_t is the price of energy. The (discretized) action space is A: an action $a_t \in A$ specifies how much energy should be charged or discharged from the storage. Prices and loads are assumed to be deterministic, in order to reduce the numerical complexity of the algorithm and to avoid the curse of dimensionality, whereas b_t is stochastic, but, given s_t, b_{t+1} is completely determined. The reward function $r_t(s_t, a_t)$ depends on both the state and the action. We will consider two forms of reward functions: the first one takes into

account only energy losses (w_t) in the network; the second one contains the losses, but multiplied by the price of the energy.

$$r_t(s_t, a_t) = -w_t \tag{1}$$

or

$$r_t(s_t, a_t) = -z_t w_t. \tag{2}$$

Then, for the period from 8 am to 4 pm, subdivided into $N-1(=32)$ 15-minute time intervals, an optimal finite-horizon policy is readily obtained from the backward induction algorithm: for every epoch t and every state s_t, the algorithm provides a function which maps the state into the action space, so that the expected total reward for the entire decision-making horizon is maximal.

This optimal policy must be connected to the optimal policy during the dark hours, that is, from 4 pm to 8 am the next day in our case. As there are no stochastic sources in this period, we have used deterministic dynamic programming, starting at 7:45 am and proceeding backwards until 4 pm the previous day, moving across $M-1(=64)$ 15-minute time intervals.

In a mathematical form, the above amounts to maximizing an expected total reward

$$\mathbb{E}_{s_1}\left\{\sum_{t=1}^{N-1} r_t(s_t, a_t)\right\} + \sum_{t=1}^{M-1} \tilde{r}_t(\tilde{s}_t, \tilde{a}_t) + \tilde{r}_M \tag{3}$$

for any initial state s_1, with respect to the sequence of actions $(a_1, \ldots, a_{N-1}, \tilde{a}_1, \ldots, \tilde{a}_{M-1})$, where the tilde accent identifies the quantities subject to no stochastic input, and $\tilde{r}_M$ is the constant terminal reward at 8 am the next day. Every action in the action sequence is constrained in two ways: first, by an upper and a lower SOC limit and by an upper limit on the current flowing from and to the storage; second, by the admissibility of the power flow. Details are given in Section III.

A. Formalization of the Optimization Problem

The state s_t is aleatory but observable and is function of clearness, SOC of the storage device and power requested by loads:

$$s_t = (c_t, b_t, \gamma_t). \tag{4}$$

Let $a_t \in A_{s_t}$ be the action to be chosen at epoch t. As previously stated, a revenue $r_t = f(s_t, a_t)$ is associated to every action a_t from state s_t.

Within every period the variables are assumed to be constant. Under this hypothesis at epoch t the revenue can be calculated.

The action a_t makes the system go from state s_t to state s_{t+1} and is chosen according to a transition kernel made up of four terms:

- $p_t(c_{t+1}|c_t)$ the probability that there is a certain clearness c_{t+1}, given that the clearness at epoch t was c_t;
- $p_t(b_{t+1}|a_t, s_t)$ the probability that the storage system reaches the state of charge b_{t+1}, given that the starting state is s_t and the the control action is a_t;
- $p_t(\gamma_{t+1}|\gamma_t)$ the probability that the load has a certain value γ_{t+1}, given that the load at epoch t was γ_t;
- $p_t(z_{t+1}|z_t)$ the probability that the energy price is z_{t+1}, given that the price at epoch t was z_t.

Thus, the transition kernel, which is the probability that the system goes from state s_t to state s_{t+1} due to the application of action a_t, can be written, in the most general form, as:

$$\begin{aligned} p_t(s_{t+1}|s_t, a_t) &= p_t(c_{t+1}|c_t)\, p_t(b_{t+1}|a_t, c_t, b_t, \gamma_t) \\ &\quad p_t(\gamma_{t+1}|\gamma_t)\, p_t(z_{t+1}|z_t) \\ &= p_t(c_{t+1}|c_t)\, p_t(b_{t+1}|a_t, s_t) \\ &\quad p_t(\gamma_{t+1}|\gamma_t)\, p_t(z_{t+1}|z_t). \end{aligned} \tag{5}$$

It should be noted that b_{t+1} is obtained as a deterministic function of (s_t, a_t), although the sequence b_t itself is stochastic. In fact, $p_t(b_{t+1}|a_t, s_t)$ is one if b_{t+1} is the SOC implied by action a_t and zero elsewhere. The load γ_t and the price z_t are assumed to be deterministic sequences. Thus, under these hypotheses the transition kernel is only influenced by the clearness. Compared to [14], although in our case decision epochs are at 15-minute intervals, we operate under similar assumptions: the market price is not random, the fixed load takes the place of committed wind power, the solar power takes the place of the actual wind power.

A control policy $\pi \in \Pi$ is made up of the decision functions $d_t : S \mapsto A$, $\pi = (d_1, \ldots, d_{N-1})$, where S is the state space and A is the action space. A decision function d_t indicates, for every system state s_t, the value of the action a_t to be taken.

The overall value of the control policy $u_t^\pi(s_t)$ is the random variable representing the expected revenue obtained by adopting control policy π for a finite-horizon problem starting at epoch t with state s_t and ending at epoch N. The aim is to find $u_1^\pi(s_1)$. This is the so-called value-to-go, which represents the expected reward of the chosen control policy π, given that the initial state is s_1. It can be derived by applying the backward recursive formula, for $t = N-1, \ldots, 1$

$$\begin{aligned} u_t^\pi(s_t) &= r_t(s_t, d_t(s_t)) + \\ &\quad + \mathbb{E}\{u_{t+1}^\pi(s_{t+1})\,|s_t, d_t\} \\ &= r_t(s_t, d_t(s_t)) + \\ &\quad + \sum_{s_{t+1}\in S} p_t(s_{t+1}|s_t, d_t(s_t))\, u_{t+1}^\pi(s_{t+1}), \\ u_N^\pi(s_N) &= r_N(s_N). \end{aligned} \tag{6}$$

The reward $r_N(s_N)$ is the value-to-go resulting from the sequence of actions taken during the dark hours.

B. Action Space

The energy stored in the battery at epoch t is $E_t = b_t E$. If control action a_t is taken, at epoch $t+1$ the total amount of energy in the storage system is

$$E_{t+1} = E_t + a_t E = (b_t + a_t) E \tag{7}$$

Deterministic dynamic programming

$\tilde{r}_M \leftarrow$ const.
$\tilde{u}^*_M(i) \leftarrow \tilde{r}_M, \forall i$
set PV production to 0 $\forall t$
for $t = M-1$ **to** 1 **do**
  update network data at epoch t
  for $i = 1$ **to** n_{SOC} states **do**
    for $j = 1$ **to** n_{SOC} states **do**
      evaluate $\Delta\tilde{E}^{ij}_t$, i.e., the energy drawn from or given to the energy storage device during the transitions from i-th SOC at t to j-th SOC at time $t+1$
      evaluate $\tilde{a}^{ij}_t$ in (7)
      chk = check feasibility of transition $i \to j$
      if chk is **false then**
        $\tilde{r}^{ij}_t := \tilde{r}_t(i, \tilde{a}^{ij}_t) \leftarrow -\infty$
      else
        evaluate $\tilde{r}^{ij}_t > -\infty$
      end if
    end for
    evaluate $\tilde{u}^*_t(i) = \max_j \left(\tilde{u}^*_{t+1}(j) + \tilde{r}^{ij}_t\right)$
    evaluate $\left(\tilde{a}^i_t\right)^*$ associated with $\tilde{u}^*_t(i)$ and $\left(\Delta\tilde{E}^i_t\right)^*$
  end for
end for

Fig. 1. Pseudo-code describing the deterministic dynamic programming algorithm implemented for the presented application

Thus, the amount of energy drawn from, or supplied to, the storage system is

$$a_t = \frac{\Delta E_t}{E}, \tag{8}$$

which can be positive or negative and is bounded by the physical and dynamical limits of the storage system. In general, these limits depend on the current state of charge, $a_t \in \left[a^{\min}(b_t),\ a^{\max}(b_t)\right]$.

C. Algorithm

The optimization algorithm is divided in two stages: a deterministic one, which runs during night-time hours, when there is no PV production, and a stochastic stage, which runs when there is PV production.

1) Deterministic: Recall that $M-1$ is the total number of "night-time" epochs, i.e., those hours when radiation is not sufficient to produce a significant amount of energy. We represented this situation by setting the clearness to zero from 4 pm to 7:45 am the next day, thus obtaining 64 quarter-of-an-hour decision epoch. Then, starting from 8 am—which is the terminal time, indexed by M—the algorithm goes back to 4 pm. The pseudo-code is shown in Fig. 1, where the tilde is used to distinguish the revenue, the exchanged energy, the actions and the value-to-go of the deterministic part of the algorithm.

2) Stochastic: For $t = N$ the optimal value-to-go is evaluated for every possible state s_N

$$u^*_N(s_N) = r^*_N(s_N), \quad \forall s_N \in S. \tag{9}$$

For t from $N-1$ to 1 the current epoch index is decreased by one unit and the optimal value of $u_t(s_t)$ is evaluated:

Stochastic dynamic programming

for $t = N-1$ **to** 1 **do**
  update network data at epoch t
  for $i = 1$ **to** n_{SOC} states **do**
    for $k = 1$ to n_{clr} clearnesses **do**
      update PV production data for clearness k
      for $j = 1$ **to** n_{SOC} states **do**
        evaluate ΔE^{ij}_t, i.e., the energy drawn from or given to the energy storage device during the transitions from i-th SOC at t to j-th SOC at time $t+1$
        evaluate a^{ij}_t in (7)
        chk = check feasibility of transition $i \to j$
        if chk is **false then**
          $r^{ikj}_t := r_t\left((i,k), a^{ij}_t\right) \leftarrow -\infty$
        else
          evaluate $r^{ikj}_t > -\infty$
        end if
        $w^{ij}_k \leftarrow 0$
        if $t+1 = N$ **then**
          $h = 1$ {at epoch N night-time hours period starts}
          $u^*_{t+1}(j,h) \leftarrow \tilde{u}^*_1(j)$
          $w^{ij}_k \leftarrow p(h|k)\, u^*_{t+1}(j,h)$ $\{p(h|k) = 1, \forall k\}$
        else
          for $h = 1$ to n_{clr} clearnesses **do**
            $w^{ij}_k \leftarrow w^{ij}_k + p(h|k)\, u^*_{t+1}(j,h)$
          end for
        end if
      end for
      evaluate $u^*_t(i,k) = \max_j \left(r^{ikj}_t + w^{ij}_k\right)$
      evaluate $\left(a^i_t\right)^*$ associated with $u^*_t(i,k)$ and $\left(\Delta E^i_t\right)^*$
    end for
  end for
end for

Fig. 2. Pseudo-code describing the stochastic dynamic programming algorithm implemented for the presented application.

$$u^*_t(s_t) = \max_{a \in A_{s_t}} \left\{ r_t(s_t, a) + \right.$$
$$\left. + \sum_{s_{t+1} \in S} p_t(s_{t+1}|s_t, a)\, u^*_{t+1}(s_{t+1}) \right\}. \tag{10}$$

In general, there can be more than one value of a maximizing (10): these values are collected in the set $A^*_{s_t,t}$. When $t = 1$ the value $u^*_1(s_1)$—i.e., the value-to-go associated with an optimal control policy π^*—is obtained. In this case π^* is a set of sequences of decision functions such as $(d^*_1, \ldots, d^*_{N-1})$, where any d^*_t selects a particular action from $A^*_{s_t,t}$.

The pseudo-code of the stochastic part of the algorithm is shown in Fig. 2.

III. Case Study

The test network used for the simulations is the European medium voltage (MV) benchmark network of the CIGRÉ [17], the single line diagram of which is depicted in Fig. 3. The network is made up of two feeders, framed by dashed lines in Fig. 3, both operating at 20 kV and fed by the high voltage grid by means of two separate 25 MVA-rated transformers. The network is made up of 15 lines and 15 nodes and feeds 14 loads, both residential and commercial. It is worth noting that the feeders can be connected by closing switch S1, thus transforming the radial distribution configuration in a meshed one. In the present simulations all switches (S1, S2, and S3) are

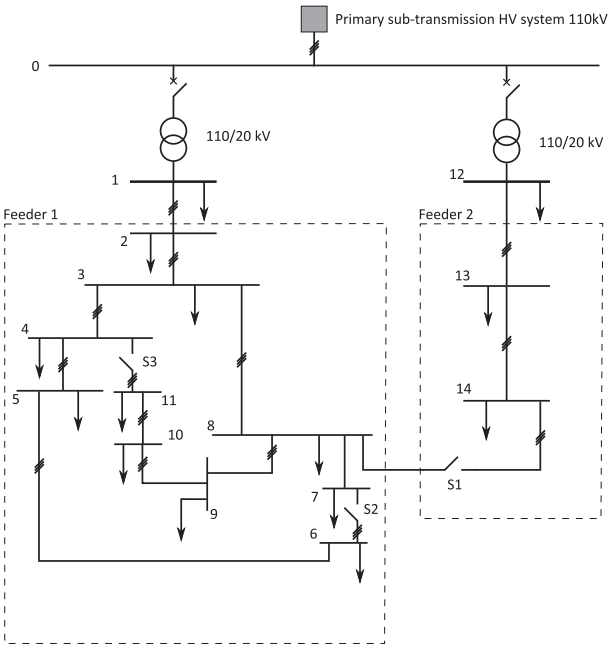


Fig. 3. Single line diagram of the MV feeders used for the simulation.

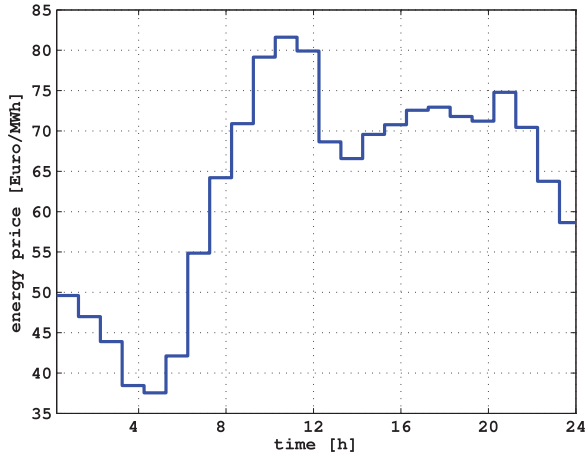


Fig. 4. Mean energy price used in the simulation and derived from Italian Power Exchange (IPEX) in 2010 [3].

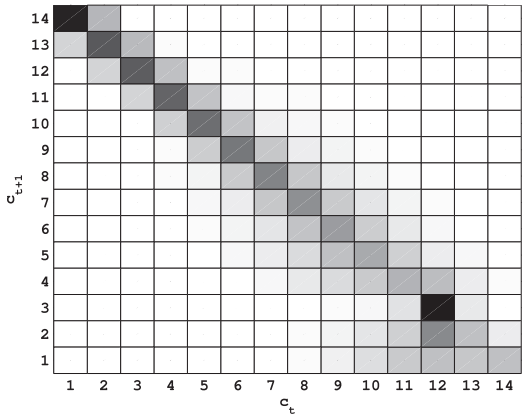


Fig. 5. Graphical representation of the transition matrix for the clearness, as reported in (11).

supposed to be closed. This configuration would guarantee, by itself, a more balanced sharing of the power flows in the grid and is one of the few structural countermeasures that a distribution system operator can exploit to avoid new installations when coping with the growing impact of renewable energy sources. For more detailed information concerning network parameters and configuration refer to [17]. The network has been implemented using MATPOWER [18]. The power flow algorithm provided by this toolbox has been used.

The price used in the simulations is depicted in Fig. 4 and is derived from the so-called PUN, the hourly-based Italian Power Exchange clearing price as it comes from the day-ahead market closure, in 2010. As reported in [3], the price profile has been built using the median values of energy price for each hour and can be considered a paradigm for energy price.

The clearness has been discretized to $n_{\text{clr}} = 14$ classes. The transition matrix for the clearness is reported in (11), shown at the bottom of the page, where the generic entry (i, j) is $p_t\,(j|i)$.

It is worth noting that each row of (11) sums up to 1. A graphical representation of the transition matrix is reported in Fig. 5, where white color is associated to 0 and black to 0.749, the maximum value of the transition matrix (11). The stationary distribution associated to the transition matrix for the clearness is plotted in Fig. 6.

$$\begin{pmatrix}
72.8 & 25.1 & 1.10 & 0.4 & 0.3 & 0.2 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
14.1 & 59.3 & 23.5 & 1.40 & 0.7 & 0.5 & 0.3 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.4 & 14.6 & 58.1 & 22.1 & 2.0 & 1.3 & 0.9 & 0.4 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.1 & 0.6 & 15.2 & 55.9 & 20.9 & 3.2 & 2.1 & 1.30 & 0.5 & 0.2 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.2 & 1.0 & 15.8 & 52.6 & 20.1 & 4.7 & 3.10 & 1.6 & 0.6 & 0.2 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.1 & 0.5 & 1.90 & 16.6 & 48.3 & 19.7 & 6.40 & 3.9 & 1.8 & 0.6 & 0.1 & 0.0 & 0.0 \\
0.0 & 0.1 & 0.3 & 1.20 & 3.6 & 17.6 & 43.5 & 19.4 & 7.8 & 4.2 & 1.7 & 0.5 & 0.1 & 0.0 \\
0.0 & 0.0 & 0.2 & 0.7 & 2.4 & 6.0 & 19.1 & 38.9 & 18.7 & 8.2 & 4.0 & 1.4 & 0.4 & 0.1 \\
0.0 & 0.0 & 0.1 & 0.3 & 1.4 & 4.1 & 9.1 & 20.8 & 34.5 & 17.3 & 7.7 & 3.3 & 1.0 & 0.3 \\
0.0 & 0.0 & 0.0 & 0.1 & 0.7 & 2.5 & 6.5 & 12.4 & 22.2 & 30.3 & 15.3 & 6.5 & 2.5 & 0.9 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.2 & 1.0 & 3.3 & 7.6 & 12.7 & 18.7 & 26.3 & 22.5 & 5.9 & 1.8 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.3 & 0.9 & 1.9 & 2.9 & 10.2 & 74.9 & 7.8 & 0.8 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 1.90 & 4.6 & 8.2 & 15.6 & 41.8 & 21.4 & 5.9 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.4 & 1.80 & 5.3 & 11.5 & 18.2 & 21.6 & 19.7 & 21.5
\end{pmatrix} \times 10^{-2} \tag{11}$$

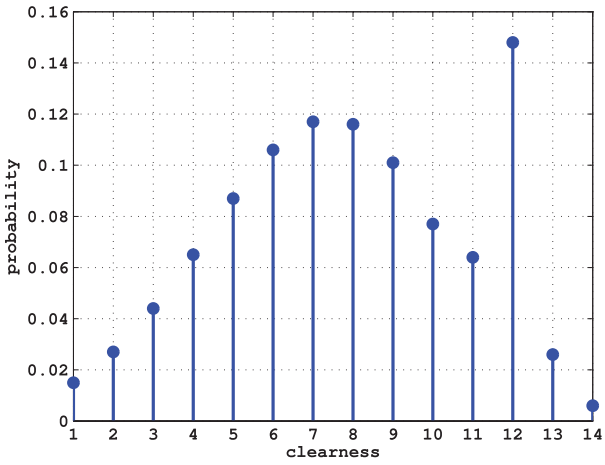


Fig. 6. Stationary distribution of the clearness.

TABLE I
LAYOUT OF THE DIFFERENT ENERGY STORAGE SYSTEM SIZES (ESSSS) USED IN THE SIMULATIONS

No.	Energy	Power
1	250 kWh	125 kW
2		250 kW
3	500 kWh	250 kW
4		500 kW
5	1 MWh	500 kW
6		1 MW

A 100 kWp-rated photovoltaic field is installed at bus 14. This bus has been selected because Feeder 2 has a low domestic-load density and can easily accommodate a large stand-alone solar field.

The energy storage device has been modeled as a combination of parallel and series connections of 50 Ah, 3.2 V-rated cells, in order to obtain the desired overall capacity. SOC has been limited to span from 0.2 to 1 and has been discretized into $n_{\mathrm{SOC}} = 81$ steps, each of which corresponds to a 1% variation of the SOC. The feasibility of a transition from one SOC to another is checked on the current involved in the transition. If the current is lower than the rated current the transition is feasible, if higher it is classified as unfeasible. For transitions in the range of SOCs 0.8–1 the limit current is derated by a 25% factor.

Different storage sizes–both energy and power—and different placements have been simulated. In particular, three energy sizes (i.e., 250 kWh, 500 kWh, and 1 MWh) have been tested. For each of these sizes two power ratings (power corresponding to full capacity and to half capacity) have been applied. A layout of the different energy storage system sizes (ESSSs) is reported in Tab. I. Moreover, each of these ESSSs have been connected to every node (with the only exception on bus 2, which, despite the graphical representation, does not supply any load).

IV. RESULTS

For every ESSS the optimal strategy (OPT) has been computed and also a non-optimal solution has been derived, so that one can appreciate the maximum amount of improvement obtained from optimization. The non-optimal solution is the worst policy (WRS), which minimizes the reward function. In order to prove the effectiveness of the proposed algorithm, 5000 days of operation have been simulated, starting every day at 8 am. These days have been generated from the transition kernel given in (11); the initial clearness was drawn from the stationary distribution of the Markov chain. Along with the optimal and the worst policy, another policy (RND) has been derived. It has been obtained through a random selection of the action among the feasible ones.

The values of losses and cost reduction (or increase) are shown in Tables II and III, respectively. In these tables the mean values of percentage deviations of optimal, random and worst policies with respect to the base case, i.e., when no storage (NOSTG) is installed in the network, are reported. Values are computed according to the equation

TABLE II
MEAN VALUES (IN PERCENTAGE) OF LOSSES REDUCTION (NEGATIVE VALUES) OR INCREASE (POSITIVE VALUES) OVER 5000 DAYS SIMULATION FOR OPTIMAL (OPT), RANDOM (RND) AND WORST (WRS) POLICIES USING A REVENUE BASED ONLY ON LOSSES MINIMIZATION. MEAN VALUE OF LOSSES WHEN NO STORAGE DEVICE IS INSTALLED IS 5.63 MWH. MAXIMUM STANDARD DEVIATIONS ARE: OPT = 0.019, RND = 0.696, AND WRS = 0.133

ESSS no.						Node	Algo	Node	ESSS no.					
1	2	3	4	5	6				1	2	3	4	5	6
-0.029	-0.030	-0.052	-0.054	-0.083	-0.084		opt		-0.502	-0.530	-0.920	-0.953	-1.560	-1.594
0.018	0.032	0.046	0.099	0.124	0.325	1	rnd	9	0.285	0.449	0.673	1.294	1.717	4.034
0.054	0.094	0.134	0.244	0.310	0.781		wrs		0.783	1.339	1.879	3.405	4.428	9.558
-0.484	-0.511	-0.889	-0.921	-1.514	-1.547		opt		-0.509	-0.536	-0.932	-0.965	-1.578	-1.611
0.273	0.434	0.637	1.216	1.622	3.781	3	rnd	10	0.291	0.461	0.684	1.319	1.759	4.110
0.750	1.280	1.798	3.222	4.237	9.050		wrs		0.794	1.358	1.907	3.466	4.499	9.990
-0.505	-0.532	-0.928	-0.961	-1.577	-1.611		opt		-0.509	-0.536	-0.931	-0.965	-1.580	-1.613
0.284	0.450	0.672	1.289	1.698	3.990	4	rnd	11	0.289	0.458	0.681	1.304	1.731	4.088
0.784	1.338	1.883	3.414	4.434	9.518		wrs		0.792	1.354	1.903	3.459	4.487	9.940
-0.514	-0.540	-0.940	-0.972	-1.588	-1.620		opt		-0.004	-0.004	-0.005	-0.005	-0.005	-0.005
0.293	0.469	0.693	1.338	1.766	4.192	5	rnd	12	0.008	0.021	0.024	0.076	0.082	0.280
0.803	1.372	1.925	3.563	4.551	10.167		wrs		0.018	0.049	0.044	0.216	0.169	0.847
-0.514	-0.540	-0.936	-0.967	-1.569	-1.603		opt		-0.243	-0.254	-0.430	-0.442	-0.667	-0.674
0.301	0.482	0.709	1.392	1.818	4.401	6	rnd	13	0.165	0.295	0.415	0.915	1.134	3.017
0.811	1.386	1.967	3.641	4.574	10.673		wrs		0.470	0.807	1.150	2.183	2.797	7.276
-0.511	-0.538	-0.932	-0.962	-1.562	-1.595		opt		-0.407	-0.427	-0.736	-0.761	-1.217	-1.240
0.298	0.476	0.702	1.380	1.819	4.371	7	rnd	14	0.246	0.408	0.594	1.197	1.547	3.833
0.807	1.380	1.959	3.625	4.554	10.626		wrs		0.660	1.204	1.741	3.047	3.820	9.211
-0.497	-0.525	-0.914	-0.947	-1.556	-1.589		opt							
0.282	0.442	0.656	1.256	1.676	3.916	8	rnd							
0.773	1.321	1.851	3.344	4.369	9.335		wrs							

TABLE III
MEAN VALUES (IN PERCENTAGE) OF COST REDUCTION (NEGATIVE VALUES) OR INCREASE (POSITIVE VALUES) OVER 5000 DAYS SIMULATION FOR OPTIMAL (OPT), RANDOM (RND) AND WORST (WRS) POLICIES USING A REVENUE BASED ON ENERGY PRICES. MEAN VALUE OF REVENUE WHEN NO STORAGE DEVICE IS INSTALLED IS €393.66. MAXIMUM STANDARD DEVIATIONS ARE: OPT = 0.018, RND = 0.709, WRS = 0.148

ESSS no.						Node	Algo	Node	ESSS no.					
1	2	3	4	5	6				1	2	3	4	5	6
-0.012	-0.013	-0.019	-0.020	-0.041	-0.044		opt		-0.219	-0.231	-0.368	-0.402	-0.565	-0.629
0.019	0.031	0.046	0.094	0.121	0.303	1	rnd	9	0.293	0.444	0.679	1.241	1.680	3.785
0.050	0.076	0.118	0.223	0.281	0.761		wrs		0.757	1.142	1.761	2.840	4.171	9.189
-0.211	-0.224	-0.358	-0.393	-0.547	-0.603		opt		-0.222	-0.235	-0.368	-0.409	-0.594	-0.659
0.279	0.423	0.649	1.172	1.599	3.580	3	rnd	10	0.295	0.452	0.687	1.273	1.731	3.882
0.724	1.109	1.682	2.700	3.984	8.686		wrs		0.767	1.180	1.784	2.886	4.232	9.464
-0.220	-0.234	-0.371	-0.410	-0.588	-0.650		opt		-0.223	-0.235	-0.368	-0.409	-0.598	-0.662
0.291	0.445	0.678	1.231	1.682	3.762	4	rnd	11	0.297	0.448	0.686	1.262	1.711	3.839
0.756	1.160	1.757	2.827	4.160	9.176		wrs		0.765	1.176	1.779	2.874	4.223	9.416
-0.223	-0.236	-0.369	-0.415	-0.614	-0.684		opt		-0.002	-0.002	-0.003	-0.003	-0.003	-0.003
0.300	0.453	0.691	1.289	1.754	3.953	5	rnd	12	0.007	0.019	0.023	0.070	0.076	0.255
0.774	1.192	1.800	2.919	4.274	9.684		wrs		0.014	0.048	0.049	0.190	0.174	0.790
-0.222	-0.237	-0.360	-0.403	-0.626	-0.701		opt		-0.103	-0.105	-0.137	-0.156	-0.328	-0.357
0.303	0.471	0.710	1.339	1.781	4.122	6	rnd	13	0.168	0.286	0.409	0.864	1.091	2.796
0.784	1.193	1.786	2.992	4.360	10.447		wrs		0.444	0.685	1.059	1.927	2.427	7.176
-0.221	-0.236	-0.358	-0.402	-0.621	-0.696		opt		-0.176	-0.180	-0.271	-0.296	-0.479	-0.544
0.306	0.471	0.706	1.333	1.790	4.109	7	rnd	14	0.250	0.398	0.598	1.151	1.511	3.594
0.781	1.188	1.776	2.980	4.343	10.370		wrs		0.633	1.020	1.602	2.583	3.642	9.048
-0.217	-0.228	-0.369	-0.403	-0.560	-0.618		opt							
0.291	0.437	0.671	1.208	1.650	3.703	8	rnd							
0.747	1.145	1.736	2.788	4.108	8.928		wrs							

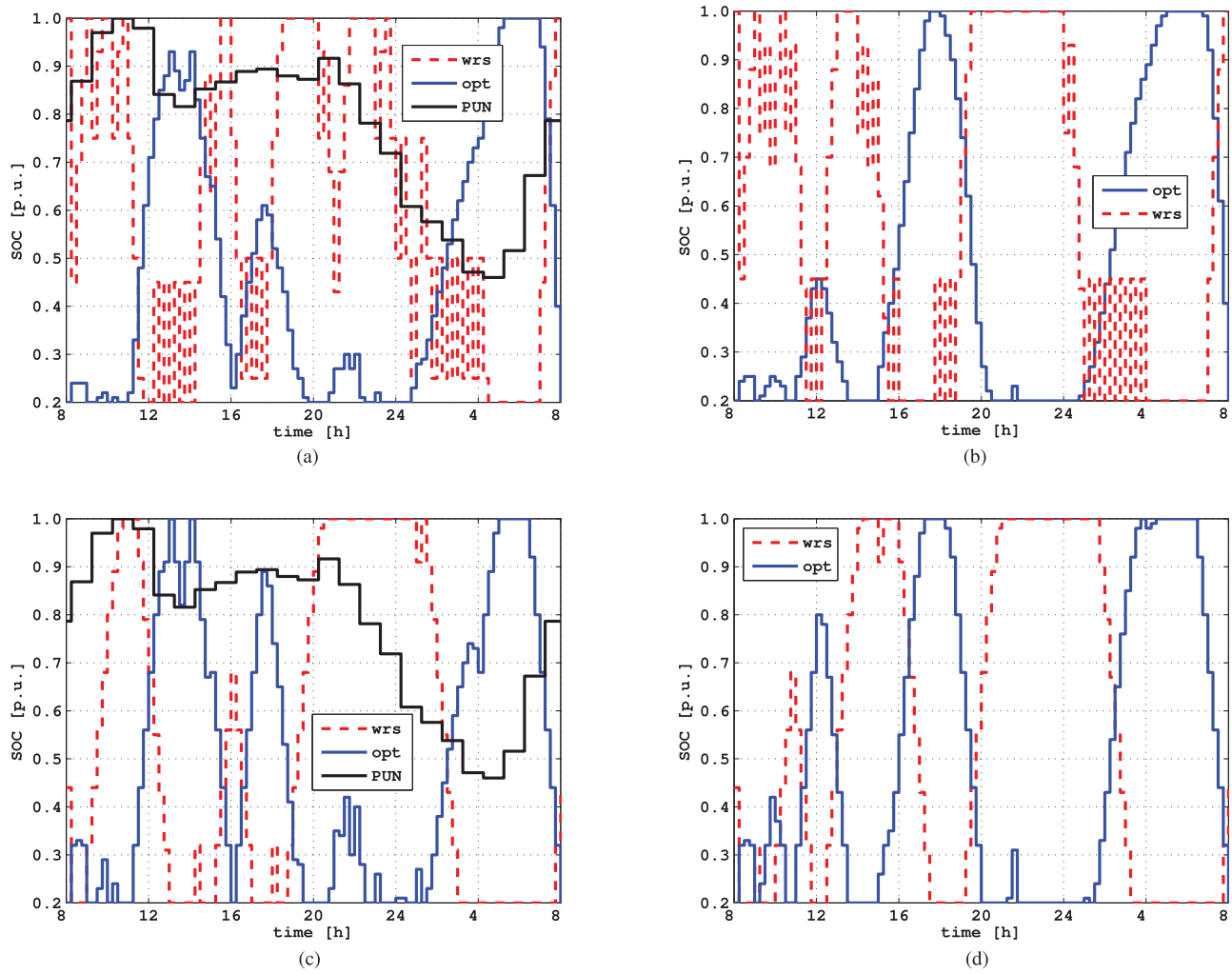


Fig. 7. SOC paths for one day of the simulations. In each plot the optimal policy (opt) is drawn with a solid blue line while the worst policy (wrs) is drawn with a dashed red line. The solid black line in the left-hand side plots is the normalized PUN, superimposed in order to highlight the influence of energy price on the policies. Top row: ESSS no. 6 = (1 MWh, 1 MW); (a) revenue based on energy price, (b) revenue based on losses minimization. Bottom row: ESSS no. 1 = (250 kWh, 125 kW); (c) revenue based on energy price, (d) revenue based on losses minimization.

$$\bar{y}^{\text{algo}} = \frac{1}{5000} \sum_{k=1}^{5000} \left(\frac{\left(y_k^{\text{algo}} - y_k^{\text{nostg}} \right) \times 100}{y_k^{\text{nostg}}} \right),$$
$$\text{algo} \in \{\text{opt}, \text{rnd}, \text{wrs}\}.$$

Maximum values of the standard deviations for each policy are reported in the captions of the tables. From the analysis of the results, it emerged that, in this particular case, the differences between best and worst policies in terms of revenue—although the values of the revenue reached by optimal policy are strictly higher than those obtained by the worst policy—are not particularly striking. This fact is due to some concurrent causes: 1) the overlap between load and PV production, 2) the small size of the energy storage device with respect to the overall consumption, and 3) the fact that energy price is high during daylight hours, when PV production is high. Moreover, the meshed grid, by itself, reduces energy losses.

The most significant results are obtained when using a losses reduction revenue function. The beneficial effect coming from a correct usage of the storage system increase when going farther from the HV/MV transformers and when using a high-capacity and high-power storage system. Losses reduction comes to a maximum figure of 1.62%. This value can be regarded as significant, as reported in [19, footnote no. 48, p. 27]. It is verified that the random policy always lays in between the optimal and the worst policy, thus confirming that every other possible management strategy would lead to no better results than the optimal one. It can be also noticed that only the optimal policy produces a reduction, while random and worst policies do not only produce no beneficial effect but can even increase both losses and costs.

Despite this little difference in the revenue, the optimal and the worst policy are extremely different. In fact, the latter tends to be the complement of the former. In order to highlight these differences, SOC paths during the same day, but under different hypotheses, are shown in Fig. 7. The day chosen is the no. 832, which corresponds to the day with the lowest number of SOC cycle of the worst algorithm for ESSS no. 6. The clearness associated to this day is reported in Fig. 8.

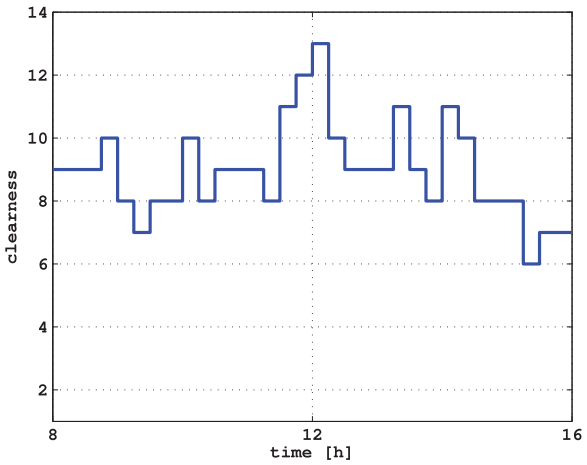


Fig. 8. Clearness during day no. 832.

TABLE IV
MEAN OF SOC CYCLES OVER 5000 DAYS SIMULATION FOR OPTIMAL (OPT), RANDOM (RND) AND WORST (WRS) POLICIES USING A REVENUE BASED ONLY ON LOSSES MINIMIZATION. MAXIMUM STANDARD DEVIATIONS ARE: OPT = 0.052, RND = 0.459, WRS = 0.270

ESSS no.						Node	Algo	Node	ESSS no.					
1	2	3	4	5	6				1	2	3	4	5	6
3.17	3.34	2.65	2.65	2.33	2.33		opt		3.36	3.66	2.90	2.92	2.46	2.46
3.67	6.93	3.67	6.93	3.67	6.93	1	rnd	9	3.67	6.93	3.67	6.93	3.66	6.95
3.55	6.10	3.55	7.46	3.24	9.66		wrs		3.11	5.34	3.30	6.67	3.24	8.28
3.38	3.69	2.91	2.95	2.48	2.48		opt		3.35	3.66	2.89	2.91	2.46	2.46
3.67	6.94	3.67	6.92	3.67	6.92	3	rnd	10	3.67	6.93	3.67	6.93	3.68	6.93
3.10	5.36	3.30	6.39	3.25	8.26		wrs		3.11	5.34	3.31	6.66	3.24	8.62
3.39	3.70	2.91	2.93	2.47	2.47		opt		3.37	3.67	2.89	2.91	2.46	2.46
3.67	6.92	3.66	6.95	3.68	6.92	4	rnd	11	3.67	6.93	3.67	6.92	3.67	6.93
3.11	5.34	3.32	6.72	3.25	8.27		wrs		3.11	5.34	3.31	6.66	3.24	8.60
3.36	3.66	2.86	2.88	2.45	2.45		opt		1.64	1.64	1.22	1.24	0.62	0.62
3.67	6.93	3.67	6.92	3.67	6.93	5	rnd	12	3.67	6.93	3.67	6.94	3.67	6.93
3.11	5.34	3.28	6.97	3.24	8.63		wrs		3.32	10.25	3.54	13.49	5.26	14.26
3.34	3.64	2.83	2.84	2.44	2.44		opt		3.04	3.13	2.55	2.55	2.20	2.20
3.67	6.93	3.68	6.93	3.67	6.93	6	rnd	13	3.67	6.93	3.67	6.93	3.67	6.93
3.11	5.29	3.46	6.97	3.13	8.87		wrs		3.31	5.87	3.23	7.73	3.46	10.17
3.33	3.63	2.83	2.84	2.44	2.44		opt		3.24	3.50	2.75	2.75	2.37	2.37
3.67	6.93	3.67	6.93	3.67	6.92	7	rnd	14	3.67	6.93	3.67	6.93	3.67	6.94
3.11	5.29	3.46	6.97	3.13	8.87		wrs		3.23	5.36	3.55	6.90	3.07	9.21
3.38	3.69	2.91	2.94	2.47	2.47		opt							
3.67	6.94	3.67	6.93	3.67	6.93	8	rnd							
3.11	5.34	3.28	6.65	3.25	8.26		wrs							

As it can be noted in ESSS no. 6, i.e., "great size", the worst policy presents a great number of oscillations. These oscillations are not present in the optimal policy. It is worth noting that no regularization term has been added in the revenue in order to obtain this behavior.

In Tables IV and V data regarding SOC cycles are shown. SOC cycles are defined as

$$\frac{\sum_{t=1}^{N-1} |\mathrm{SOC}_{t+1} - \mathrm{SOC}_t|}{2\,(\mathrm{SOC}_{\max} - \mathrm{SOC}_{\min})}.$$

Compared to loss minimization, energy-price-based policy increases the number of cycles, in both optimal and worst policies. In addition, energy-price-based policy obtains, for small and medium ESSSs, lower standard deviation, i.e., the solution is "more stable" and less dependent on clearness. It can be noted that, when connected to node 12, i.e., the one just after the HV/MV transformer of Feeder 2, the optimal policy tends to use the energy storage device less, with a mean number of SOC cycles below 0.7 for the "great size" energy storage systems

TABLE V
MEAN OF SOC CYCLES OVER 5000 DAYS SIMULATION FOR OPTIMAL (OPT), RANDOM (RND) AND WORST (WRS) POLICIES USING A REVENUE BASED ON ENERGY PRICES. MAXIMUM STANDARD DEVIATIONS ARE: OPT = 0.042, RND = 0.458, WRS = 0.208

ESSS no.						Node	Algo	Node	ESSS no.					
1	2	3	4	5	6				1	2	3	4	5	6
3.80	4.28	3.25	3.31	2.25	2.25		opt		4.09	4.82	3.55	3.80	2.72	2.74
3.67	6.93	3.67	6.94	3.67	6.95	1	rnd	9	3.67	6.93	3.67	6.94	3.67	6.93
3.72	7.11	3.73	8.40	3.96	11.16		wrs		3.31	7.60	3.72	6.92	3.81	10.17
4.11	4.85	3.56	3.87	2.76	2.80		opt		4.07	4.81	3.54	3.82	2.68	2.70
3.67	6.93	3.67	6.93	3.67	6.94	3	rnd	10	3.67	6.93	3.67	6.93	3.68	6.94
3.31	7.60	3.72	6.92	3.76	10.18		wrs		3.35	7.60	3.72	6.92	3.81	10.24
4.11	4.85	3.57	3.84	2.74	2.77		opt		4.09	4.82	3.55	3.82	2.69	2.72
3.67	6.92	3.67	6.94	3.67	6.94	4	rnd	11	3.67	6.94	3.67	6.92	3.67	6.93
3.34	7.60	3.72	6.92	3.76	10.22		wrs		3.35	7.60	3.72	6.92	3.76	10.24
4.06	4.80	3.54	3.80	2.66	2.68		opt		1.94	1.94	1.34	1.34	0.67	0.67
3.67	6.93	3.67	6.93	3.68	6.94	5	rnd	12	3.67	6.93	3.67	6.95	3.67	6.93
3.40	7.60	3.72	6.92	3.81	10.31		wrs		3.16	11.68	5.08	13.61	6.17	14.52
4.02	4.70	3.48	3.67	2.59	2.60		opt		3.72	4.15	2.96	2.98	2.08	2.08
3.67	6.92	3.67	6.93	3.67	6.94	6	rnd	13	3.67	6.93	3.67	6.93	3.66	6.93
3.44	7.59	3.60	6.92	3.81	10.69		wrs		3.61	7.23	3.72	8.29	3.70	11.92
4.01	4.69	3.48	3.67	2.59	2.60		opt		3.95	4.55	3.32	3.39	2.43	2.43
3.67	6.94	3.68	6.93	3.67	6.93	7	rnd	14	3.68	6.94	3.67	6.93	3.67	6.94
3.44	7.60	3.58	6.92	3.81	10.65		wrs		3.38	7.27	3.71	7.24	3.81	11.05
4.11	4.86	3.56	3.86	2.75	2.78		opt							
3.67	6.94	3.67	6.92	3.67	6.93	8	rnd							
3.31	7.60	3.72	6.92	3.76	10.16		wrs							

and with both revenue functions. Also in the other ESSSs the mean value of SOC cycles is lower compared to other locations.

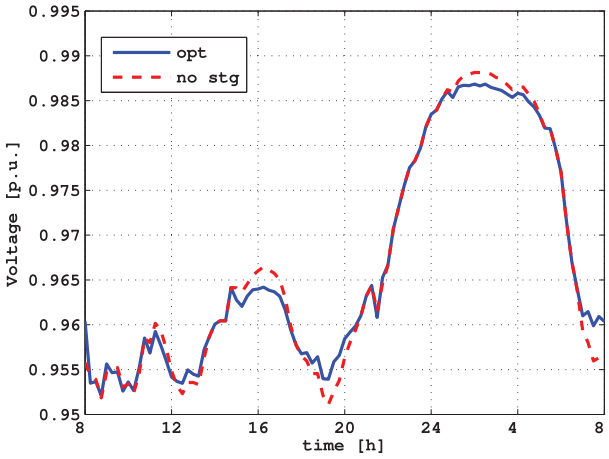
Looking at these data as a whole, some considerations can be derived. The six ESSSs can be compared clustering them into three orders of groups: 1) ESSSs with the same capacity (but different power), e.g., (1–2), (3–4), and (5–6); 2) ESSSs with the same power (but different capacity), e.g., (2–3) and (4–5); and 3) ESSSs in which rated power is half (or equal to) the rated capacity, e.g., (1–3–5) and (2–4–6). Let x_i be the mean number of SOC cycles for the i-th ESSS. From the first group, it can be noted that, when doubling the power (ESSS $i+1$), it should always be

$$x_{i+1}^{\mathrm{opt}} > x_i^{\mathrm{opt}}. \tag{12}$$

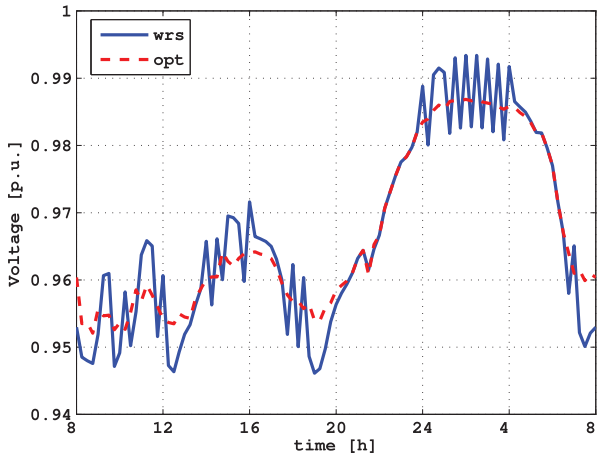
If not, it would mean that the optimal policy is not influenced by a doubling of the rated power of the converter interfacing the storage device to the grid. Thus, the same result would be attained with smaller (and, possibly, cheaper) equipment. As an example, this condition is verified in node 12 for ESSSs (1–2) and in node 14 for ESSSs (5–6). From the second set, it can be noted that, when doubling the capacity (ESSS $i+1$), it should always be

$$x_{i+1}^{\mathrm{opt}} \geq \frac{x_i^{\mathrm{opt}}}{2}. \tag{13}$$

If not, it would mean that, despite doubling the capacity of the storage device, the optimal policy would handle a smaller amount of energy, thus demonstrating that the optimal policy found for ESSS i, indeed, is not optimal. If (13) holds as an equality, the conclusion that capacity-doubling is ineffective could be inferred. As an example, this condition is verified in node 12 for ESSSs (4–5). These conclusions could be useful in determining sizing and placing of the energy storage device. From the third group some considerations can be inferred by comparing optimal and worst policies: when, moving from one



(a)



(b)

Fig. 9. Voltage profiles at bus no. 6 for day no. 3683 and ESSS no. 6 with storage system connected at bus 5. The optimization criteria is losses minimization. (a) comparison between the optimal policy (solid blue line) and the base case with no storage (dashed red line); (b) comparison between the optimal policy (dashed red line) and the worst policy (solid blue line).

ESSS to the next one, the difference between worst and optimal policy is such that

$$\left(x_{i+2}^{\text{wrs}} - x_{i+2}^{\text{opt}}\right) - \left(x_{i}^{\text{wrs}} - x_{i}^{\text{opt}}\right) > 0. \tag{14}$$

This means that there is some sort of "regularity" in the optimal policy. It can also be noted that in the set of ESSSs (1–3–5), $x_1^{\text{wrs}} - x_1^{\text{opt}}$ is, generally, negative. Thus, it can be inferred that the combination of low capacity and low power is not suited for this kind of application. The only exceptions are for node 12 in Tab. V and for nodes 12 and 13 in Tab. IV and derive from network configuration and PV plant location.

Finally, a comparison between voltage profiles is shown in Fig. 9. The voltage profiles of bus no. 6 for day no. 3683 and ESSS no. 6 with storage system connected at bus 5 (optimization only on losses). This particular ESSS, along with the placement of the energy storage system has been chosen because it gives the best result. The particular day is when the maximum spread between highest and lowest voltage occurs, while the bus selected is that which displays the minimum voltage. This selection has been made because of the huge amount of data (more than 37 million of voltage values for each optimization criterion). It can be noticed that the optimal management of the energy storage system enhance the voltage profiles.

V. Conclusion

The application of the MDP framework to the optimal scheduling of a storage device in the presence of solar generation is a tool which can contribute effectively to network management. The superiority of this method is intrinsic, because it takes into account the whole chain of consequences of every possible action across the entire decisional horizon, instead of considering only the here-and-now reward at every epoch. The Markovian assumption for the clearness stochastic process offers a simplified yet accurate enough representation of the uncertainty of the solar source, so that the optimal strategy is valid (on average) over all possible instances of the clearness over a day and it can be applied to all days with a similar law of the clearness pattern. Then, by estimating different transition matrices for every season in the year from long enough series of observed radiation data, a year-long optimal scheduling strategy for the storage can be obtained.

The specific testing environment chosen for this article has demonstrated that, although the improvement measured by the selected objective function can be modest (be it in terms of network losses or energy prices), this methodology can optimize other features of the system, such as the number of cycles of the storage, so that its life will be increased. Incidentally, the direct usage of the number of cycles as objective function within this framework would not be possible, because the optimization algorithm is based on a chain of evaluations of reward functions spanning only one epoch at a time. Then, either relying directly on the main objective function or on performance measures arising as by-products of the optimization, this methodology could be used to assess the effectiveness of a range of system layouts, such as the placement of the storage, its size and its power rating.

Further work should consider different systems, such as networks equipped with a larger solar field or multiple storages. A problem that could arise is that, after executing the policy optimization for multiple layouts, it might not be easy to discriminate among them, especially if the associated rewards are close to each other. Therefore, appropriate choice criteria of the reward function or of by-product performance measures should be studied.

Acknowledgment

The authors would like to thank the anonymous reviewers for their careful reading of the manuscript and for suggesting several modifications which greatly helped to improve the paper.

References

[1] M. Farrokhifar, S. Grillo, and E. Tironi, "Loss minimization in medium voltage distribution grids by optimal management of energy storage devices," in *Proc. IEEE PowerTech*, 2013, pp. 1–5.

[2] M. Farrokhifar, S. Grillo, and E. Tironi, "Optimal placement of energy storage devices for loss reduction in distribution networks," in *Proc. IEEE Innov. Smart Grid Technol. Eur.*, 2013, pp. 1–5.
[3] S. Grillo, M. Marinelli, S. Massucco, and F. Silvestro, "Optimal management strategy of a battery-based storage system to improve renewable energy integration in distribution networks," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 950–958, Jun. 2012.
[4] V. Marano, G. Rizzo, and F. A. Tiano, "Application of dynamic programming to the optimal management of a hybrid power plant with wind turbines, photovoltaic panels and compressed air energy storage," *Appl. Energy*, vol. 97, pp. 849–859, Sep. 2012.
[5] H. Morais, P. Kádár, P. Faria, Z. A. Vale, and H. Khodr, "Optimal scheduling of a renewable micro-grid in an isolated load area using mixed-integer linear programming," *Renew. Energy*, vol. 35, no. 1, pp. 151–156, Jan. 2010.
[6] Z. Ziadi, S. Taira, M. Oshiro, and T. Funabashi, "Optimal power scheduling for smart grids considering controllable loads and high penetration of photovoltaic generation," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2350–2359, Sep. 2014.
[7] M. L. Puterman, *Markov Decision Processes*. Hoboken, NJ, USA: Wiley, 1994.
[8] C. A. Glasbey, "Nonlinear autoregressive time series with multivariate Gaussian mixtures as marginal distributions," *Appl. Stat.*, vol. 50, pp. 143–154, 2001.
[9] M. A. Murtaza and M. Tahir, "Optimal data transmission and battery charging policies for solar powered sensor networks using Markov decision processes," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2013, pp. 992–997.
[10] J. Donadee and M. Ilić, "Stochastic optimization of grid to vehicle frequency regulation capacity bids," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 1061–1069, Mar. 2014.
[11] P. M. van de Ven, N. Hegde, L. Massoulié, and T. Salonidis, "Optimal control of end-user energy storage," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 789–797, Jun. 2013.
[12] J. Qin, R. Sevlian, D. Varodayan, and J. Rajagopal, "Optimal electric energy storage operation," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2012, pp. 1–6.
[13] N. Gast, D.-C. Tomozei, and J.-Y. Le Boudec, "Optimal generation and storage scheduling in the presence of renewable forecast uncertainties," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1328–1339, May 2014.
[14] L. A. Hannah and D. B. Dunson, "Approximate dynamic programming for storage problems," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 337–344.
[15] V. Musolino, L. Piegari, and E. Tironi, "Technical and economical evaluation of storage systems in naval applications," in *Proc. Int. Conf. Clean Electr. Power (ICCEP)*, 2013, pp. 120–127.
[16] G. Celli, S. Mocci, F. Pilo, and M. Loddo, "Optimal integration of energy storage in distribution networks," in *Proc. IEEE PowerTech Conf.*, 2009, pp. 1–7.
[17] K. Strunz *et al.*, "Benchmark systems for network integration of renewable and distributed energy resources," CIGRÉ TF C6.04.02, Tech. Rep. 575, Apr. 2014.
[18] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2014.
[19] VV. AA., "Smart Distribution System: Promozione selettiva degli investimenti nei sistemi innovativi di distribuzione di energia elettrica. Orientamenti iniziali. (Smart Distribution System: Selective promotion of investments in innovative electric distribution systems. Initial guidelines.)," Autorità per l'Energia Elettrica, il Gas e il Sistema Idrico (AEEGSI), Tech. Rep. 255/2015/R/eel, May 2015, in Italian [Online]. Available: http://www.autorita.energia.it/it/docs/dc/15/255-15.jsp

Samuele Grillo (S'05–M'09) received the Laurea degree in electronic engineering, and the Ph.D. degree in power systems from the University of Genova, Italy, in 2004 and 2008, respectively. He is currently an Assistant Professor (tenure track) with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy. His research interests include smart grids, integration of distributed renewable sources, and energy storage devices in power networks, optimization, and control techniques applied to power systems.

Antonio Pievatolo received the Ph.D. in statistics from Padua University, Padua, Italy. He is currently a Researcher with CNR-IMATI. His research interests include industrial statistics. In particular, he likes to address problems in engineering and technology using stochastic modeling tools, with a preference for Markov and point processes. His main contributions can be found in the application of stochastic modeling to reliability and to electrical power systems.

Enrico Tironi received the M.S. degree in electrical engineering from the Politecnico di Milano, Milan, Italy, in 1972. In 1972, he joined the Department of Electrical Engineering, Politecnico di Milano, where he is currently a Full Professor with the Dipartimento di Elettronica, Informazione e Bioingegneria. His research interests include power electronics, power quality, distributed generation, and energy storage systems. He is a member of the Italian Standard Authority (C.E.I.), the Italian Electrical Association, and the Italian National Research Council Group of Electrical Power System.