ESAIM: PROCEEDINGS AND SURVEYS, February 2019, Vol. 65, p. 46-67
B. Bouchard, J.-F. Chassagneux, F. Delarue, E. Gobet and J. Lelong, Editors

REGRESSION MONTE CARLO FOR MICROGRID MANAGEMENT

Clemence Alasseur[1], Alessandro Balata[2], Sahar Ben Aziza[3], Aditya Maheshwari[4], Peter Tankov[5] and Xavier Warin[6]

Abstract. We study an islanded microgrid system designed to supply a small village with the power produced by photovoltaic panels, wind turbines and a diesel generator. A battery storage system device is used to shift power from times of high renewable production to times of high demand. We build on the mathematical model introduced in [14] and optimize the diesel consumption under a "no-blackout" constraint. We introduce a methodology to solve microgrid management problem using different variants of Regression Monte Carlo algorithms and use numerical simulations to infer results about the optimal design of the grid.

1. Introduction

A Microgrid is a network of loads and energy generating units that often include renewable sources like photovoltaic (PV) panels and wind turbines alongside more traditional forms of thermal electricity production. These microgrids can be part of the main grid or isolated. Communities in rural areas of the world have long now enjoyed the installation of isolated microgrid systems that provide a reliable and often environment-friendly source of electricity to meet their power needs.

The elementary purpose of a microgrid is to provide a continuous electricity supply from the variable power produced by renewable generators while minimizing the installation and running costs. In this kind of systems, the uncertainty of both, the load and the renewable production is high and its negative effect on the system stability can be mitigated by including a battery energy storage system in the microgrid. Energy storage devices ensure power quality, including frequency and voltage regulation (see [12]) and provide backup power in case of any contingency. A dispatchable unit in the form of diesel generator is also used as a backup solution and to provide baseload power.

In this paper, we consider a traditional microgrid serving a small group of customers in islanded mode, meaning that the network is not connected to the main national grid. The system consists of an intermittent renewable generator unit, a conventional dispatchable generator, and a battery storage system. Both the load and the intermittent renewable production are stochastic, and we use a stochastic differential equation (SDE) to model directly the residual demand, that is, the difference between the load and the renewable production. We then set up a stochastic optimization problem, whose goal is to minimize the cost of using the diesel generator

[1] EDF R&D - FIME, Palaiseau, France; e-mail: clemence.alasseur@edf.fr
[2] University of Leeds, Woodhouse Lane, Leeds LS2 9JT, United Kingdom; e-mail: A.Balata@leeds.ac.uk
[3] University of Tunis El Manar,ENIT-LAMSIN, BP.37, Le Belvédère 1002 Tunis, Tunisia; e-mail: sahar.benaziza@enit.utm.tn
[4] University of California, Santa Barbara, USA; e-mail: aditya_maheshwari@umail.ucsb.edu
[5] CREST-ENSAE, Palaiseau, France; e-mail: peter.tankov@ensae.fr
[6] EDF R&D - FIME, Palaiseau, France; e-mail: xavier.warin@edf.fr

Article published online by EDP Sciences and available at https://www.esaim-proc.org or https://doi.org/10.1051/proc/201965046

plus the cost of curtailing renewable energy in case of excess production, subject to the constraint of ensuring reliable energy supply. A regression Monte Carlo method from the mathematical finance literature is used to solve this stochastic optimization problem numerically. Three variants of the regression algorithm, called grid discretization, Regress now and Regress later are proposed and compared in this paper. The numerical examples illustrate the performance of the optimal policies, provide insights on the optimal sizing of the battery, and compare the policies obtained by stochastic optimization to the industry standard, which uses deterministic policies.

The optimization problem arising from the search for a cost-effective control strategy has been extensively studied. Three recent survey papers [15, 20, 21] summarize different methods used for optimal usage, expansion and voltage control for the microgrids. Heymann et. al. [13, 14] transform the optimization problem associated with the microgrid management into an optimal control framework and solve it using the corresponding Hamilton Jacobi Bellman equation. Besides proposing an optimal strategy, the authors also compare the solution of the deterministic and stochastic representation of the problem. However, similarly to most PDE methods, this approach suffers from the curse of dimensionality and as a result, it is difficult to scale. The main contribution of this paper is to solve the microgrid control problem using Regression Monte Carlo algorithms. In contrast to existing approaches, the method used in this paper is more easily scalable and works well in moderately large dimensions [5]. It is fair to mention here that the problem we study in the following is however low dimensional as it displays one source of randomness and one degenerate controlled process.

Identifying the optimal mix, the size and the placement of different components in the microgrid is an important challenge to its large scale use. The papers [17, 18] use mixed-integer linear programming to address the design problem and test their model on a real data set from a microgrid in Alaska. In a similar work, [19] studied the economically optimal mix of PV, wind, batteries and diesel for rural areas in Nigeria. In [11], optimal battery storage sizing is deduced from the autocorrelation structure of renewable production forecast errors. In this paper, we propose an alternative approach for the optimal sizing of the battery energy storage system, assuming stochastic load dynamics and fixed lifetime of the battery. Our in-depth analysis of the system behavior leads to practical guidelines for the design and control of islanded microgrids.

Finally, several authors [7–9] used stochastic control techniques to determine optimal operation strategies for wind production – storage systems with access to energy markets. In contract to these papers, in the present study, energy prices appear only as constant penalty factors in the cost functional, and the main focus is on the stable operation of the microgrid without blackouts.

The rest of the paper is organized as follows: In section 2 we describe the microgrid model and introduce the different components of the system, in section 3 we translate the problem of managing the microgrid in a stochastic optimization problem and present the dynamic programming equation that we intend to solve numerically. Section 4 introduces the numerical algorithms used to solve the control problem, we give a general framework for solving the dynamic programming equation and we then provide three algorithms for the approximation of conditional expectations. In section 5 we illustrate the results of the numerical experiments, identify the best algorithm among those we studied and then employ it to analyze the system behavior. We conclude with section 6 where the estimated policy for the stochastic problem is compared, in an appropriate manner, with a deterministically trained one; the aim is to provide evidence that industry-widespread deterministic approaches underperform stochastic methods.

2. Model description

In this section, we will discuss the topology of the microgrid, its operation, components and their respective dynamics. Although we discuss a simplified microgrid model, more complicated typologies can be studied using straightforward generalizations of the methods presented in this paper.

Consider a microgrid serving a small, isolated village; most of the power to the village is supplied by generating units whose output has zero marginal cost, is intermittent and uncontrolled. Additional power is supplied by a controlled generator whose operations come alongside a cost for the microgrid owner (either the community itself or a power utility). Often the intermittent units include PV panels and wind turbines, while the controlled

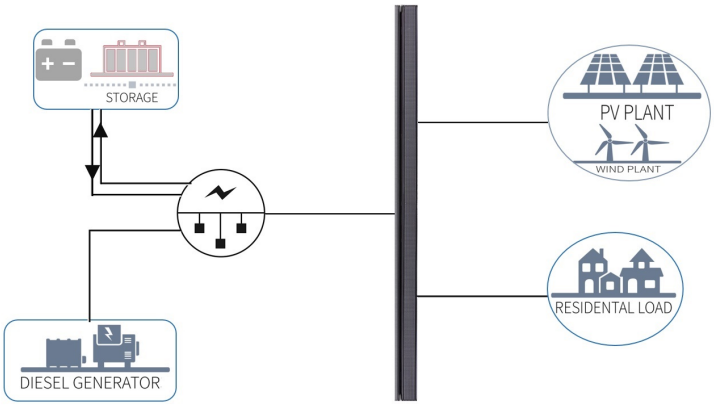


FIGURE 1. The figure above shows an example of microgrid topology that contains all the elements in our model. The network is arranged as follows: photovoltaic panels and wind turbines provide renewable generation, a diesel generator provides dispatchable power for the village and a battery storage system is used to inject or withdraw energy.

unit is often a diesel generator. In order to fully exploit the free power generated by the renewable units at times when production exceeds the demand, microgrids are equipped with energy storage devices. These can be represented by a battery energy storage system.

The introduction of the battery in the system not only allows for inter-temporal transfer of energy from times when demand is low, to times when it is higher, but also introduces an element of strategic behavior that can be employed by the system controller, to minimize the operational costs. Without an energy storage, diesel had to be run at all times demand exceeded production. When a battery is installed, intensity and timing of output from the diesel generator can be adjusted to move the level of charge of the battery towards the most cost effective levels.

In figure 1 we propose a schematic description of the system which might help the readers to familiarize themselves with the microgrid, whose components are described more in depth in the following subsections.

Remark 1. *Note that for convenience, in the following, we will work in discrete time only. This setting is not restrictive as in reality measurements of the systems are repeated at a given, finite, frequency. We also consider a finite optimization horizon represented by the number of periods over which we want to optimize the system operations indicated by T*

2.1. Residual Demand

Consider two stochastic processes L_t and R_t, the former represents the demand/load and the latter the production through the renewable generators. Notice that both processes are uncontrolled and they represent, respectively, the unconditional withdrawal and injection of power in the system (constant during time step). For the purpose of managing the microgrid, the controller is interested only in the net effect of the two processes denoted by the process X_t:

$$X_t = L_t - R_t \; ; \quad t \in \{0, 1, \ldots, T\}. \tag{1}$$

Remark 2. *The state variable X_t represents the residual demand of power at each time t, such that for $X_t > 0$, we should provide power through the battery or diesel generator and for $X_t < 0$ we can store the extra power in the battery.*

For simplicity, we model the residual demand as an AR(1) process, the discrete equivalent of an Ornstein–Uhlenbeck process. In practical applications we expect X_t to be an $\mathbb{R}$-valued mean reverting process with

many different sources of noise and time dependent random parameters; our choice of using an AR(1) avoids the cumbersome notation coming from multiple noise sources still providing scope for generalization. The process X_t is driven by the following difference equation, starting from an initial point $X_0 = x_0$:

$$X_{t+1} = X_t + b(\Lambda_t - X_t)\Delta t + \sigma\sqrt{\Delta t}\,\xi_t\,; \quad t \in \{0, 1, \dots, T\} \tag{2}$$

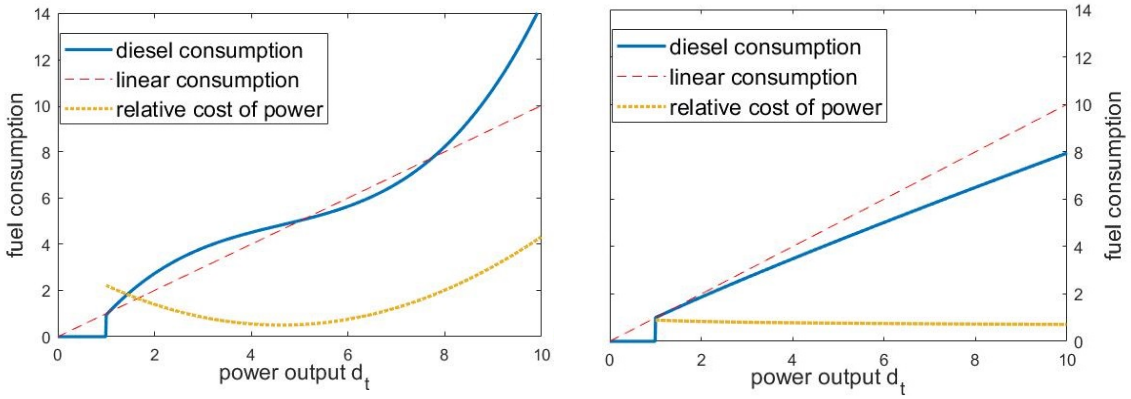
where $\xi_t \sim \mathcal{N}(0,1)$, Δt is the amount of time before new information is acquired, b is the mean reversion speed, σ the volatility of the process and Λ_t is the mean reversion level (typically deterministic function of time).

Remark 3. *In real applications the deterministic function Λ_t should represent the best forecast available for future residual demand at the time of the estimation of the policy.*

2.2. Diesel generator

The Diesel generator represents the controlled dispatchable unit. The state of the generator is represented by $m_t = \{0, 1\}$. If $m_t = 0$ then the diesel generator is OFF, while it is ON when $m_t = 1$. When the engine is ON, it produces a power output denoted by $d_t \in [d_{min}, d_{max}]$ at time t, for $d_{min} > 0$.

Notice that, in addition, when the engine is turned ON, an extra amount of fuel is burned in order for the generator to warm up and reach working regime. We model the cost of burning extra fuel with a switching cost $\mathcal{K}$ that is paid every time the switch changes from 0 to 1. The fuel consumption of the diesel generator is modeled by an increasing function $\rho(d_t)$ which maps the power d_t produced during one time step into the quantity of diesel necessary for such output. Denoting by P_t the price of fuel at time t, the cost of producing d_t KW of power at one time step is $P_t\rho(d_t)$; for simplicity we take a constant price of the fuel $P_t = p$. Two examples of efficiency functions ρ are described in figure 2.



(A) $\rho(d) = \frac{(d-6)^3+6^3+d}{10}$ (B) $\rho(d) = d^{0.9}$

FIGURE 2. The panels above show two examples of efficiency function (liters/KW), on the left $\rho(d) = \frac{(d-6)^3+6^3+d}{10}$, typical of a generator designed to operate at medium regime, on the right $\rho(d) = d^{0.9}$, typical of a generator designed to operate a full capacity.

2.3. Dynamics of the Battery

The storage device is directly connected to the microgrid and therefore its output is equal to the imbalance between demand X_t and diesel generator output d_t, when this is allowed by the physical constraint. The battery therefore is discharged in case of insufficiency of the diesel output and charged when the diesel generator and renewables provide a surplus of power.

Let us denote the state of charge of the battery at time t as I_t^d and its maximum capacity as I_{max}. If the power rating of the battery is given by $B^{\max}$ and $B^{\min}$, where $B^{\max}$ and $B^{\min}$ represent respectively the maximum and minimum output with $B^{\min} < 0 < B^{\max}$, its power output B_t^d at time t is defined as:

$$B_t^d = \frac{I_t^d - I_{\max}}{\Delta t} \vee \left(B^{\min} \vee (X_t - d_t) \wedge B^{\max}\right) \wedge \frac{I_t^d}{\Delta t}. \tag{3}$$

Intuitively, $B_t^d < 0$ refers to the charging of the battery and $B_t^d > 0$ refers to the supply of power from the battery. The inner terms in equation (3) capture the constraints due to maximum power output/input to the battery and the outer terms capture the effect of the capacity constraints on the power output of the battery. Notice then that an energy storage has a limited amount of capacity after which it can not be charged further, as well as an "empty" level below which no more power can be provided from the battery. The dynamics of the controlled process I_t^d is described by the following equation:

$$I_{t+1}^d = I_t^d - B_t^d \Delta t, \quad t \in \{0, 1, \ldots, T-1\}, \quad I_0^d = w_0 \tag{4}$$

here $I_t^d \in [0, I_{max}]$ and $B_t^d \in [B^{\min}, B^{\max}]$. For simplicity we assume that the battery is 100% efficient. Notice that we used superscript d on B^d and I^d to highlight the dependence of these processes on the controlled diesel output d_t.

Intuition tells us that the bigger the battery, the less diesel will be needed to run the operations of the microgrid. This is true because a bigger battery would allow to store for later use a bigger proportion of the excess power produced by the renewables. Batteries however are very expensive, and the cost per KWh of capacity scales almost linearly for the kind of devices we consider in this paper (parallel connection of smaller batteries), hence it is important to find the optimal size of battery for the needs of each specific microgrid.

2.4. **Management of the Microgrid**

The purpose of the microgrid is to provide a cheap and reliable source of power supply to at least match the demand. Therefore, we search for a control policy for the diesel generator which minimizes the operating cost and produces enough electricity to match the residual demand. In order to assess how well we are doing in supplying electricity, we introduce the controlled imbalance process S_t defined as follows:

$$S_t = X_t - B_t^d - d_t \quad t \in [0, T] \tag{5}$$

Ideally, the owner of the Microgrid would like to have $S_t = 0 \quad \forall\, t$. This situation represents the perfect balance of demand and generation. When $S_t > 0$ we observe a *blackout*, residual demand is greater than the production meaning that some loads are automatically disconnected from the system. The situation $S_t < 0$ is defined as a curtailment of renewable resources and takes place when we have a surplus of electricity.

We treat the two scenarios, blackout and curtailment asymmetrically. To ensure no-blackout $S_t \leqslant 0$ and regular supply of power, we impose a constraint on the set of admissible controls:

$$\begin{aligned} S_t &\leqslant 0 \\ \text{i.e. } d_t &\geqslant X_t - B_t^d. \end{aligned} \tag{6}$$

However, for $S_t < 0$ i.e. surplus of electricity, we penalize the microgrid using a proportional cost denoted by C. Large penalty would lead to low level of curtailment and can be thought of as a parameter in the subsequent optimization problem.

A rigorous mathematical description of the microgrid management problem follows in section 3.

3. Stochastic optimization problem

We state now the stochastic control problem for the diesel generator operating in a microgrid system as described in section 2. In practice we seek a control that minimizes the cost of diesel usage $p\rho(d)$, the switching cost $\mathcal{K}$ and the curtailment cost $C|S_t|\mathbb{1}_{\{S_t<0\}}$, under the no black-out constraint $S_t \leqslant 0$.

Note that, given the type of control we have on the diesel generator, we can frame the optimization problem as a special case of stochastic control problems known as optimal switching problems.

Let us denote by $(\mathcal{F}_t)_{t\geqslant 0}$ the filtration generated by the random variables $\{\xi_t\}_{t\geqslant 0}$, which represent the only randomness in the system, that is, we define $\mathcal{F}_t = \sigma(\xi_i, i < t)$ for $t \geqslant 1$, and $\mathcal{F}_0$ to be the trivial σ-field. We require the control process $(d_t)_{t\geqslant 0}$ to be adapted to this filtration or, in other words, no future information should be used to determine its value. Under this assumption, the residual demand process $(X_s)_{s=0}^t$, the state of charge process $(I_s^d)_{s=0}^t$ and the current regime m_t, become adapted to $(\mathcal{F}_t)_{t\geqslant 0}$. The objective of the controller is to minimize the following cost functional

$$\mathbb{E}\left[\sum_{s=0}^{T-1} \mathbb{1}_{\{m_s - m_{s-1}=1\}}\mathcal{K} + p\rho(d_s) + C|S_s|\mathbb{1}_{\{S_s<0\}} + g(I_T^d)\right],$$

with the convention $m_{-1} = 0$, where g is a terminal condition which might be linked with situations where the battery has been rented and has to be returned with the same level of charge otherwise a penalty might be applied. The minimization is carried out over the set of admissible strategies $\mathcal{U}$, containing all $(\mathcal{F}_t)_{t\geqslant 0}$-adapted controls $(d_t)_{t\geqslant 0}$ such that

$$d_t \geqslant X_t - B_t^d \quad \forall t \tag{7}$$

$$d_t \in [d_{min}, d_{max}] \cup \{0\}. \tag{8}$$

$$B_t^d = \frac{I_t^d - I_{\max}}{\Delta t} \vee \left(B^{\min} \vee (X_t - d_t) \wedge B^{\max}\right) \wedge \frac{I_t^d}{\Delta t} \tag{9}$$

where (7) represents the no-blackout constraints translated for the power produced by the diesel generator, (8) represents the minimum and maximum power output of the generator and (9) models the physical constraints of the battery: maximum input/output power and maximum capacity.

Since the state dynamics is Markovian, the optimal control is of feedback type and can be computed using the dynamic programming approach (see [3, Chapter 8]). To formulate this approach, we define the pathwise value $\mathcal{J}_t$ starting from time t, given by

$$\mathcal{J}_t = \sum_{s=t}^{T-1} \mathbb{1}_{\{m_s - m_{s-1}=1\}}\mathcal{K} + p\rho(d_s) + C|S_s|\mathbb{1}_{\{S_s<0\}} + g(I_T^d). \tag{10}$$

The value function is then defined as follows.

$$V(t,x,w,m) = \min_{d\in\mathcal{U}_t}\left\{\mathbb{E}\left[\mathcal{J}_t \middle| X_t = x, I_t^d = w, m_{t-1} = m\right]\right\}, \tag{11}$$

where the class $\mathcal{U}_t$ contains admissible controls "starting from time t": processes $(d_s)_{s=t}^{T-1}$ adapted to the filtration $\mathcal{F}_s^t := \sigma(\xi_u, t \leqslant u < s$ and satisfying the constraints (7), (8) and (9) between t and $T-1$.

The dynamic programming principle associated to (11), decomposes the problem on a single interval into two optimal control problems: an optimal switching problem between being in the regime ON or OFF, and another absolutely continuous control problem assuming the regime is ON. The equation reads as follows:

$$V(t,x,w,m) = \min_{d}\Big(\mathbb{1}_{\{\mathbf{1}_{d\neq 0}-m=1\}}\mathcal{K} + p\rho(d) + C|S|\mathbb{1}_{\{S<0\}} + \mathcal{C}(t,x,w;d)\Big), \tag{12}$$

$$\text{subject to} \quad d \geqslant x - B, \quad d \in [d_{min}, d_{max}] \cup \{0\}, \tag{13}$$

$$\text{where} \quad B = \frac{w - I_{\max}}{\Delta t} \vee \left(B^{\min} \vee (x-d) \wedge B^{\max}\right) \wedge \frac{w}{\Delta t} \quad \text{and} \quad S = x - B - d \tag{14}$$

$$\text{and} \quad \mathcal{C}(t,x,w;d) = \mathbb{E}[V(t+1, X_{t+1}, I_{t+1}, \mathbf{1}_{d\neq 0})|X_t = x, I_t = w]. \tag{15}$$

In order to ensure that the set of admissible controls is nonempty we introduce the following assumption:

Assumption 1. *The diesel generator is powerful enough to supply demand at all times, i.e there is always a control d that satisfies the blackout constraint.*

Remark 4. *We enforce assumption 1 by redefining the residual demand process with a truncated version of (1), such that $\tilde{X}_t = \min(X_t, X_{\max})$ is the residual demand. In practice this is reasonable because the maximum power that could be required from the microgrid is known apriori and the diesel generator is generally sized to the maximum capacity installed on the system. For the sake of notational simplicity, we will drop the $\sim$ on the variable $\tilde{X}_t$ from the following sections.*

Note that (12) provides a direct technique to solve problem (11), iterating backward in time from a known terminal condition and solving a static, one period, optimization problem at each time step. The only difficulty in this procedure lies in the estimation of conditional expectations of future value function, which can not be computed exactly. In the next section 4 we will focus on the numerical solution of (11).

4. Numerical Resolution

In this section we describe the algorithm which we want to employ in the solution of the energy management problem for the Microgrid system described in section 3. The main mathematical difficulty comes from the approximation of conditional expectations in (12), which we will tackle using a family of methods called Regression Monte Carlo. For our purposes we assume that the one step optimization problem can be solved either by extensive search, or by any more efficient method preferred by the reader. Here we discretize the set of possible controls into a finite collection, as a result the optimization is straightforward.

The algorithm we propose fully exploits the dynamic programming formulation (12): we start generating a set of simulations (scenarios) of the process X, which we will refer to as training points, then we optimize our policy so that it performs well, on average (weighted on the probability of each scenario), on the different scenarios.

In practice, we initialize the value function at last time step in the backward procedure to be equal to the terminal condition g. We then iterate backward in time and at each time step over each training point we choose the control that minimizes the sum of one step cost function and the estimated conditional expectation of the future costs $\tilde{\mathcal{C}}(t,x,w,m;d)$. Note that, as expected, the conditional expectation is a function of time, the state of the system (x,w) and the state of the diesel generator, represented by the ON/OFF switch m and the control d.

As the iteration reaches the initial time point we collect a set of optimal actions for each time step and many different scenarios; in addition, since the problem is Markovian, we can summarize such strategies in the form of control maps: best action at each time t given a pair of state variables (X_t, I_t) and state of the diesel generator m_t. We propose three different techniques to compute $\tilde{\mathcal{C}}$ in section 4.1.

A fair assessment of the quality of the control policies approximated by the algorithm just introduced is obtained by running a number of forward Monte Carlo simulations of the residual demand, controlling the system using such policies and then taking the average performance.

We give a general description of the pseudo code in algorithm 1.

Remark 5. *Notice that it is typical of Regression Monte Carlo algorithms to provide the optimal policy only implicitly, in the form of minimizer of an explicit parameterized function. The outputs of the algorithm are therefore the parameters (regression coefficients) of such function.*

Algorithm 1 Regression Monte Carlo algorithm for Microgrid management

---

input: number of basis K, number of training points M, discretisation of the inventory D, time-steps N.

optimization:
if Inventory discretisation **then**
  Generate a customary grid $\{w_0, \dots, w_D\}$ points over the domain of I_t.
  Simulate $\{X_t^j\}_{j,t=1}^{M',N}$ according to its dynamics where $M' = M/(D+1)$;
  Define $\{X_t^j, I_t^j\}_{j=1}^{M}$ as cross product of $\{X_t^j\}_{j=1}^{M'}$ and $\{w_j\}_{j=0}^{D}$ for $\forall t$
if Regression 2D **then**
  if Regress Later **then**
    Generate $\{X_t^j, I_t^j\}_{j,t=1}^{M,N}$ accordingly to a distribution μ;
  if Regress Now **then**
    Generate $\{X_t^j\}_{j,t=1}^{M,N}$ according to its dynamics and $\{I_t^j\}_{j,t=1}^{M,N}$ according to a distribution μ;
Initialize the value function $V(N, X_N^j, I_N^j, 1) = V(N, X_N^j, I_N^j, 0) = g(I_N^j), \quad \forall j = 1, \dots, M$;
for $t = N$ to 1 **do**
  Compute the approximated continuation value $\tilde{\mathcal{C}}$ using Algorithms 3 or 2
  for $j = 1$ to M **do**
    for $m = 0$ to 1 **do**
      $F = \tilde{\mathcal{C}}(X_t^j, I_t^j; 0, 0)$

$$V(t, X_t^j, I_t^j, m) = \begin{cases} \left(\min\limits_{d \in \mathcal{U}_t \setminus \{0\}} \left\{ p\rho(d) + C|S_t| \mathbb{1}_{\{S_t<0\}} + \tilde{\mathcal{C}}(X_t^j, I_t^j; 1, d) \right\} + \mathcal{K} \mathbb{1}_{\{m=0\}} \right) \wedge F & \text{if } 0 \in \mathcal{U}_t \\ \min\limits_{d \in \mathcal{U}_t} \left\{ p\rho(d) + C|S_t| \mathbb{1}_{\{S_t<0\}} + \tilde{\mathcal{C}}(X_t^j, I_t^j; 1, d) \right\} + \mathcal{K} \mathbb{1}_{\{m=0\}} & \text{otherwise} \end{cases}$$

simulation:
initialize processes
for $t = 1$ to $N-1$ **do**
  for $j = 1$ to M **do**
    $F_1 = \tilde{\mathcal{C}}(X_t^j, I_t^j; 0, 0)$
    $F_2 = \min\limits_{d \in \mathcal{U}_t \setminus \{0\}} \left\{ p\rho(d) + C|S_t| \mathbb{1}_{\{S_t<0\}} + \tilde{\mathcal{C}}(X_t^j, I_t^j; 1, d) \right\} + \mathcal{K} \mathbb{1}_{\{m_t^j=0\}}$
    $m_{t+1}^j = \mathbb{1}_{\{(0 \notin \mathcal{U}_t) \text{ or } (0 \in \mathcal{U}_t \text{ and } F_2 < F_1)\}}$
    if $m_{t+1}^j = 1$ **then**
      $d_t = \operatorname{argmin}\limits_{d \in \mathcal{U}_t} \left\{ p\rho(d) + C|S_t| \mathbb{1}_{\{S_t<0\}} + \tilde{\mathcal{C}}(X_t^j, I_t^j; 1, d) \right\}$
    compute X_{t+1}^j and $I_{t+1}^j = I_t^j - B_t^d \Delta t$
    $J_{t+1}^j = J_t^j + p\rho(d_t) + C|S_t| \mathbb{1}_{\{S_t<0\}} + \mathcal{K} \mathbb{1}_{\{m_{t+1} - m_t = 1\}}$
$V(0, x, w, m) = \frac{1}{M} \sum_{j=1}^{M} (J_N^j + g(I_N^j))$

output: control policy $\{d_t\}$, value function V.

4.1. **Regression for continuation value**

In this section we present the numerical techniques we use to estimate conditional expectations $\mathcal{C}(t,x,w,m;d)$ in algorithm 1. These techniques belong to the realm of Regression Monte Carlo methods, and in particular these specifications allow to deal with degenerate controlled processes (the inventory). We focus on two main variants: a two dimensional approximation of the conditional expectation and a discretisation technique which considers a collection of one dimensional approximations.

In particular, we test three algorithms: Grid Discretisation, Regress Now and Regress Later. Grid Discretization is characterized by a one dimensional projection in the residual demand dimension repeated at different inventory points. Regress Now/Later, on the other hand, use a two dimensional regression in residual demand and inventory. Moreover, while Grid Discretization and Regress Now require projection of the value function at $t+1$ on $\mathcal{F}_t$ measurable basis functions, Regress Later requires an $\mathcal{F}_{t+1}$ projection. For details on these techniques see [2] for regress later, [4, 22] for GD and [6] for 2D regress now. Note that in the three algorithms we repeat the regression approximation for both values of m. An open source platform has also been developed to numerically solve wide variety of stochastic optimization problems in [10].

Let us denote by $\{X_t^j\}_{j=1}^M$ the collection of training points at time t, similar notation is used for the inventory $\{I_t^j\}_{j=1}^M$.

4.1.1. *Grid Discretisation*

Grid discretisation is characterized by a one dimensional approximation of the conditional expectation repeated at different levels of inventory. Let $\Upsilon_I = \{w_0 = 0, \dots, w_D = I_{max}\}$ be a discretisation of the state space of the inventory and $\{X_t^j\}_{j=1,t=1}^{M,N}$ be generated from a forward simulation of the dynamics of X. We define the approximation of the continuation value on the grid Υ_I by regressing the set of value functions $\{V(t+1, X_{t+1}^j, w_i)\}_{j=1}^M$ over the basis functions $\{\phi_k(x)\}_{k=1}^K$ for each $\{w_i\}_{i=0}^D$, obtaining:

$$\hat{\mathcal{C}}(t,x,w_i;m) = \sum_{k=1}^{K} \alpha_{k,i,m}^t \phi_k(x)\,, \quad i = 0,\, 1,\, \dots,\, D,$$

where we compute a collection of regression coefficients through least square minimization

$$\boldsymbol{\alpha}_{i,m}^t = \underset{a\in\mathbb{R}^K}{\operatorname{argmin}} \Big\{ \frac{1}{M} \sum_{j=1}^{M} \big(V(t+1, X_{t+1}^j, w_i, m) - \sum_{k=1}^{K} a_k \phi(X_t^j) \big)^2 \Big\},$$

where we define $\mathbb{R}^K \ni \boldsymbol{\alpha}_{i,m}^t = (\alpha_{1,i,m}^t,\, \dots,\, \alpha_{K,i,m}^t)$.

Note that the least square projection is a sample estimation of the L^2 projection induced by the conditional expectation, for this reason we can approximate the function $\mathcal{C}(t,\cdot)$ using a least square projection of the value function at time $t+1$.However, as we have not included the inventory in the basis functions, we need to interpolate between values of $\hat{\mathcal{C}}(t,x,w_i;m)$ in order to obtain an estimation of the value function for $I_t \in (w_i, w_{i+1})$. Let us define by $\tilde{\mathcal{C}}(t,x,w;m,d)$ the linear interpolation

$$\tilde{\mathcal{C}}(t,x,w;m,d) = \omega(t,w,d)\hat{\mathcal{C}}(t,x,w_i,m) + \big(1-\omega(t,w,d)\big)\hat{\mathcal{C}}(t,x,w_{i+1},m)\,, \quad w - B_t^d \Delta t \in [w_i, w_{i+1}),$$

where $\omega(t,w,d) = \frac{w_{i+1} - w + B_t^d \Delta t}{w_{i+1} - w_i}$ and $i = 0,\, \dots,\, D$.
Details of the algorithms are given in the pseudocode 2.

4.1.2. *2D Regression*

Contrary to the grid discretisation approach, the 2D regression methods approximate the conditional expectation of the value function as a surface, function of both residual demand X and inventory I, without the need for interpolation. In the problem we consider, the control only acts on a degenerate process and we can

Algorithm 2 Regression technique for continuation value: Grid Discretisation

input: $\{V(t+1, X^j_{t+1}, I^j_{t+1}, m)\}_{j=1}^M$, $\{\phi_k\}_{k=1}^K$.

1: **for** $i = 0$ to D **do**

2: $\quad \boldsymbol{\alpha}^t_m = \underset{a}{\operatorname{argmin}} \Big\{ \sum_{j=1}^{M} \Big(V(t+1, X^j_{t+1}, w_i, m) - \sum_{k=1}^{K} a_k \phi_k(X^j_t) \Big)^2 \Big\}$;

3: $\quad$ Define $\hat{\mathfrak{C}}(t, x, w_i, m) = \sum_{k=1}^K \alpha^t_{k,i,m} \phi_k(x)$, $m = 0, 1$;

4: Define $\tilde{\mathfrak{C}}(t, x, w; m, d) = \frac{w_{i+1} - w + B^d_t \Delta t}{w_{i+1} - w_i} \hat{\mathfrak{C}}(t, x, w_i; m, d) + \frac{w - B^d_t - w_i}{w_{i+1} - w_i} \hat{\mathfrak{C}}(t, x, w_{i+1}; m, d)$, $w \in [w_i, w_{i+1})$, $m = 0, 1$.

output: $\tilde{\mathfrak{C}}$, $\{\alpha^t_{k,i,m}\}_{k=1,i=1,m=0}^{K,D,1}$.

therefore test two specifications of the method: "Regress Now", where we project over $\{\phi_k(X_t, I_{t+1})\}_{k=1}^K$ and "Regress Later", where we project over $\{\phi_k(X_{t+1}, I_{t+1})\}_{k=1}^K$. The terminology Regress Now or Regress Later is attributed to the time step of the exogenous variable X_t used in the projection.

In Regress Now, we generate training points $\{X^j_t\}_{j=1,t=1}^{M,N}$ from a forward simulation of the dynamics of X and $\{I^j_t\}_{j=1,t=1}^{M,N}$ from a distribution μ_N on $[0, I_{max}]$ independently. In Regress Later, on the other hand, we generate both processes $\{X^j_t, I^j_t\}_{j=1,t=1}^{M,N}$ from an appropriate distribution μ_L, for details see [2]. Notice that we do not require to have any dependence between the different time steps in Regress Later, contrary to the exogenous dimension in Regress Now where the samples $\{X^j_{t+1}|X_t\}_{j=1}^M$ are simulated using the empirical distribution of the process $X_{t+1}|X_t$. In the following we will generalize the discussion of the two approaches by using the subscript r with realization t to indicate Regress Now algorithm and $t+1$ to indicate Regress Later. As training measures we choose μ_N to be the Lebesgue measure on $[0, I_{\max}]$ and μ_L to be Lesbegue measure on $[0, I_{\max}] \times [-X_{\max}, X_{\max}]$.

Remark 6. *Although in this paper we chose μ_N to be independent of the dynamics of $(X_t)_{t \geqslant 0}$, in a parallel work [16] authors have discussed in detail the effect of the correlation between $\{X^j_t\}_{j=1}^M$ and $\{I^j_{t+1}\}_{j=1}^M$ on the performance of Regress Now algorithm.*

The regression coefficients in the 2D regression Monte Carlo method are computed by least-square projection as:

$$\boldsymbol{\alpha}^t_m = \underset{a \in \mathbb{R}^K}{\operatorname{argmin}} \Big\{ \mathbb{E}\Big[\big(V(t+1, X^j_{t+1}, I^j_{t+1}, m) - \sum_{k=1}^{K} a_k \phi(X^j_r, I^j_{t+1}) \big)^2 \Big] \Big\},$$

where we define $\mathbb{R}^K \ni \boldsymbol{\alpha}^t_m = (\alpha^t_{1,m}, \ldots, \alpha^t_{K,m})$.

Let us recall, denoting by $\boldsymbol{\phi}$ the vector $\big(\phi_1(\cdot), \ldots, \phi_K(\cdot)\big)$, that the coefficients $\boldsymbol{\alpha}^t_m$ can be computed explicitly by

$$\boldsymbol{\alpha}^t_m = \Big(\mathbb{E}_\mu\big[\boldsymbol{\phi} \boldsymbol{\phi}^T \big] \Big)^{-1} \mathbb{E}_\mu \Big[V(t+1, X_{t+1}, I_{t+1}, m) \boldsymbol{\phi} \Big]^T \approx \Big(\sum_{j=1}^{M} \boldsymbol{\phi} \boldsymbol{\phi}^T \Big)^{-1} \sum_{j=1}^{M} V(t+1, X^j_{t+1}, I^j_{t+1}, m) \boldsymbol{\phi}^T$$

and therefore, even though the regression coefficients are random (sample average approximation of expectations with respect to the measure μ) they are independent of $\mathcal{F}_t$. Given the previous remark we can estimate the conditional expectation of future value through:

$$\begin{aligned} \tilde{\mathfrak{C}}(t, x, w; m, d) &= \mathbb{E}\Big[\sum_{k=1}^{K} \alpha^t_{k,m} \phi_k(X_r, I_{t+1}) \Big| X_t = x,\, I_t = w,\, d_t = d \Big] \\ &= \sum_{k=1}^{K} \alpha^t_{k,m} \mathbb{E}\Big[\phi_k(X_r, I_{t+1}) \Big| X_t = x,\, I_t = w,\, d_t = d \Big]. \end{aligned}$$

The explicit value of $\mathbb{E}\Big[\phi_k(X_r, I_{t+1})\Big|X_t = x,\, I_t = w,\, d_t = d\Big]$ now depends on r, i.e. whether we are using "Regress Now" or "Regress Later" to deal with the uncontrolled residual demand. In the first case we simply obtain, from the measurability of X_t,

$$\mathbb{E}\Big[\phi_k(X_t, I_{t+1})\Big|\mathcal{F}_t\Big] = \phi_k(x, w - B_t^d\Delta t) =: \tilde{\phi}_k(x, w, d).$$

In the second case we need to compute the expectation with respect to the randomness contained in the transition function from X_t to X_{t+1} and we simply write

$$\mathbb{E}\Big[\phi_k(X_{t+1}, I_{t+1})\Big|\mathcal{F}_t\Big] = \mathbb{E}_\xi\Big[\phi_k(x + b(\Lambda_t - x)\Delta t + \sigma\sqrt{\Delta t}\xi, w - B_t^d\Delta t)\Big] =: \hat{\phi}_k(x, w, d).$$

Remark 7. *For polynomial basis functions, i.e.* $\phi_k(X_{t+1}, I_{t+1}) := X_{t+1}^p I_{t+1}^q$, *the conditional expectation* $\hat{\phi}_k(x, w, d)$ *can be written in closed form as:*

$$\begin{aligned}\hat{\phi}_k(x, w, d) &= \mathbb{E}\big[X_{t+1}^p, I_{t+1}^q\big|X_t = x, I_t = w, d_t = d\big] \\ &= I_{t+1}^q \sigma^p dt^{\frac{p}{2}} \sum_{k=0}^{p} \mathbb{I}_{\{(p-k)\ is\ odd\}} \binom{p}{k} \Big(x\frac{1-\lambda dt}{\sigma\sqrt{dt}}\Big)^k \prod_{j=1}^{\frac{p-k}{2}} (2j-1)\end{aligned}$$

Using the notation just introduced we can summarize the differences between the two techniques in the following table:

	ϕ_k	$\mathbb{E}[\phi_k\|X_t, I_t, d_t]$	$\mathcal{C}(t, x, w, m; d)$
RN	(X_t, I_{t+1})	$\phi_k(X_t, I_t - B_t^d\Delta t)$	$\sum_{k=1}^K \alpha_{k,m}^t \tilde{\phi}_k(x, w, d)$
RL	(X_{t+1}, I_{t+1})	$\mathbb{E}[\phi_k(X_t + b(\Lambda_t - X_t)\Delta t + \sigma\sqrt{\Delta t}\xi, I_t - B_t^d\Delta t)]$	$\sum_{k=1}^K \alpha_{k,m}^t \hat{\phi}_k(x, w, d)$

Details of the algorithms are given in the pseudocode 3 .

Algorithm 3 Regression technique for continuation value: 2D Regression

input: $\{V(t+1, X_{t+1}^j, I_{t+1}^j, m)\}_{j=1}^M$, $\{\phi_k\}_{k=1}^K$.

1: **if** Regress Later **then**
2: $\quad r = t + 1$
3: **else if** Regress Now **then**
4: $\quad r = t$
5: $\boldsymbol{\alpha}_m^t = \underset{a}{\operatorname{argmin}} \Big\{ \sum_{j=1}^M \Big(V(t+1, X_r^j, I_{t+1}^j, m) - \sum_{k=1}^K a_k \phi_k(X_r^j, I_{t+1}^j)\Big)^2 \Big\},\ m = 0, 1;$
6: Define $\tilde{\mathcal{C}}(t, x, w; m, d) = \sum_{k=1}^K \alpha_{k,m}^t \mathbb{E}[\phi_k(X_r, I_{t+1})|x, w, d]$

output: $\tilde{\mathcal{C}}$, $\{\alpha_{k,m}^t\}_{k=1,m=0}^{K,1}$.

Remark 8. *A natural question that comes to mind when reading this section is how the error bound on the estimation of conditional expectations influences the estimation of the value function. To answer this question we suggest the reader to check the proof of convergence in [1] where the authors compute error bounds in the case of control of fully random Markov processes for the Regress Later scheme. In the Regress Now case some studies of the convergence behaviour are available but are relevant only for stopping and switching problems, where the choice of the control does not influences the future distribution of the state variables. The inventory problem case that we are treating in this paper however does not fall into the case studies in [1] because of the singularity of the transition density of the inventory process that falls outside their assumptions.*

5. Numerical Experiments

In this section we use the algorithms introduced in section 4 to solve a simple instance of the microgrid management problem. We fix some base parameters and test the three algorithms; the one performing best is then used to study the sensitivity of the control policy and of the operational costs on changes in system parameters, hoping to gain some insight on the optimal design of the microgrid.

We now list the base parameters chosen for the numerical experiments; notice that the "s" column indicates whether a sensitivity analysis is run for such parameter. For the meaning of the parameters refer to section 2.

parameter	value	s
T	$100h$	
Δt	$0.25h$	
b	0.5	*
σ	2	*
Λ_t	0, $\forall t$	

parameter	value	s
I_{max}	10 KWh	*
$\rho(d)$	$\frac{(d-d^*)^3+(d^*)^3+d}{10}\frac{litre}{\text{KW}}$	
d^*	6 KW	
p	1€	
$g(i)$	0, $\forall i$	

parameter	value	s
d_{min}	1KW	
d_{max}	10KW	
K	5€	*
C	0€	*

According to the parameters table above, and recalling remark 4 the residual demand has the following dynamics:

$$X_{t+1} = \left(X_t(1 - 0.5\Delta t) + \sigma\sqrt{\Delta t}\xi_t\right) \wedge 10, \quad t \in \{0, 1, \dots, T-1\}, \tag{16}$$

where $\xi_t \sim \mathcal{N}(0, 1)$.

We decided to use such simple dynamics for illustrative purposes in order to make the sensitivity of the optimal control policy to the remaining parameters more straight forward to understand.

Consider now that for the parameters listed above, the problem is time homogeneous. We have also observed empirically that the estimated continuation values tend to forget the terminal condition rather quickly. We show in Figure 3 that the regression coefficients for all algorithms converge to a stationary value time steps, suggesting that optimization ran for longer time horizons would not bring any noticeable effect to control policy. Since all three methods use polynomial basis of degree two for the projection, it also allows for easy comparison of the dynamics of the coefficients across methods. For example, at inventory level $I = 0$ the dynamics of the coefficient for x achieves same stationary level for both Grid Discretization and Regress Now. Although an exact comparison is not possible between Regress Now and Regress Later, we continue to observe similar sign and dynamics for each of the coefficients. However, getting away with almost no noise in the dynamics of the estimated coefficients of Regress Later compared to Regress Now is essentially magical.

As a result, we define a stationary policy $d(x, w, m)$ to be used in a longer time horizon than the one employed for its estimation which performance are comparable to the time dependent policy $d(t, x, w, m)$.

We finally tested the value of both stationary and time dependent policy and found that the performance of the stationary policy is comparable to that of the time dependent policy.

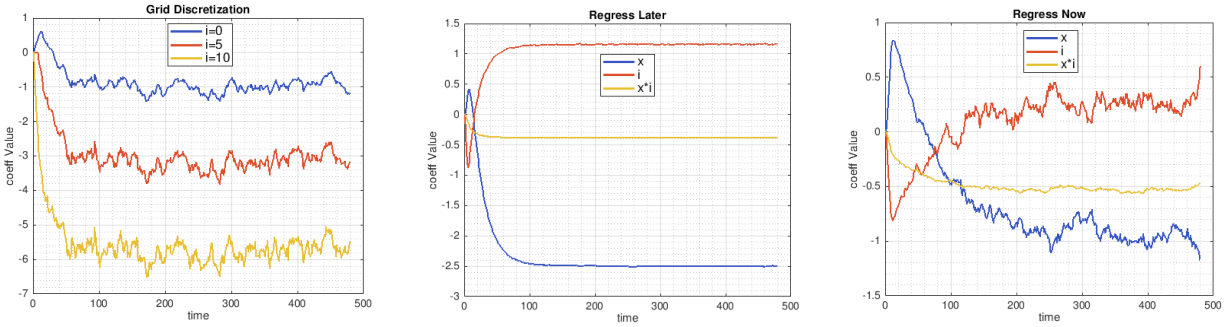
5.1. **Analysis of the controllers**

In this section we compare the control policies estimated by the three algorithms and we try to assess whether one of the approaches is preferable.

5.1.1. *Control maps*

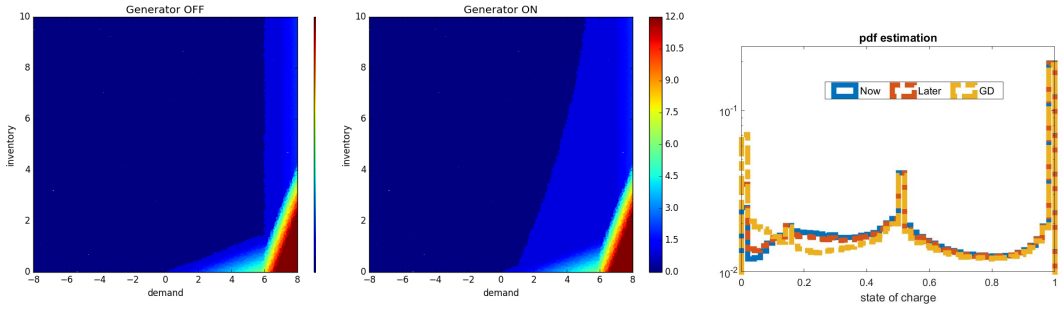
We compare now the stationary control policies produced by the different algorithms; recall that these policies are feedback to the state, i.e. can be written as function $d_m(x, w)$. Figure 4 displays an example of the feedback control policy in the form of control map, a graphical representation of the value of the optimal control for each pair (x, w).

We observed that the three policies agree with the intuition that the diesel generator should produce more power when residual demand is high and inventory is low. We can also notice that the switching cost influences the policy, forcing the diesel to keep running for longer in order to charge the battery sufficiently and avoid



(A) (B) (C)

FIGURE 3. In the three panels above we display the estimated regression coefficients corresponding to the basis $\{x, i, x\,i\}$ in the case of 2D regression, and $\{x\}$ at three different inventory levels for GD for $m_t = 1$. Although we used basis function up to polynomial degree 2, we present few coefficients for clarity of presentation. Notice that the time axis is inverted to show the number of time steps computed backward. Remarkable smooth coefficients are computed by the Regress Later algorithm.



(A) (B) (C)

FIGURE 4. In the figure above we show, in the two left-most panels, an example of control map produced by the Regress Later algorithm at time $t = 0$. Notice the difference depending on the state of the generator. In the right-most panel we display the estimated probability density function of the state of charge of the battery associated with the use of the three policies. It can be observed that Regress Later and Grid discretization induce very similar distributions.

turning ON and OFF the generator too often. Just by observation of the control maps little difference can be found among the algorithms, we display in Figure 4 the effect of the control policy on a the state of charge of the battery. It can be observed from the estimated unconditional probability density of the process I that the policies induced by Regress Now and Regress Later are very similar. Both seem to induce a peculiar mass of probability around $I_n = 2.5$, differentiating the behavior of the inventory compared to Grid Discretization. The distribution of the state of charge, obtained by plotting the histogram of all simulations over all time steps, shows that Regress Now and Regress Later does not fully exploit the whole inventory but rather they are more conservative, saving energy to avoid to turn ON the diesel generator in the future. In the next section we will investigate the value associated to this control maps.

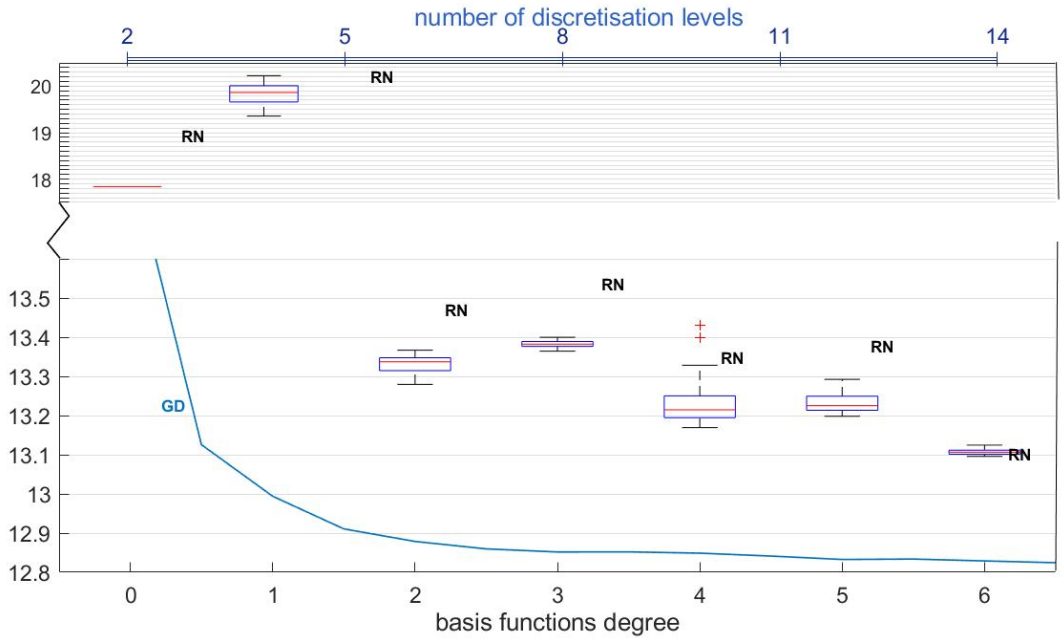


FIGURE 5. The figure in display shows the reduction in average operating cost (value function for given initial condition) when higher degree polynomials are added to the basis functions, in the case of RN regression, or more inventory points in Grid Discretisation (see upper tick bar, where we indicate the number of discretisation levels). Notice the peculiar behaviour of even/odd degree of basis functions in the RN regressions. Similar analysis was performed for Regress Later and the results are available on request.

5.1.2. *Performance of the policies*

In order to assess the performance of each policy in an unbiased manner, we select a collection of simulated paths of the residual demand process X, and record the costs associated with managing the microgrid as indicated by each control map.

We first study how the quality of each policy improves when we increase the computational budget M (and the complexity of the projection K) for each algorithm to compute the stationary policy. In Figure 5, we show the estimated value of the policy when the initial state of the system is $(x, i, m) = (0, 5, 0)$ for polynomial basis functions of increasing degree, for 2D regression. In case of GD we increase the number of discretisation points for the inventory. In particular we make the computational time increase by providing the problem with more training points and more parameters to use in the definition of $\mathcal{C}$ as increasing the number of basis functions. In the case of 2D regression, surprisingly, we notice that the performance of the estimated control improves only when polynomials of even degree are added, and the effect is more prominent for Regress Later.

We notice from the comparison that Grid Discretisation converges quickly, resulting in the best algorithm in terms of trade off between running time and precision. Among the 2D regressions, we observe similar bias for Regress Now and Regress Later (not displayed in order to maintain clear presentation, but available on request), however latter has lower standard error. This is not surprising because Regress Later has only one element of approximation error due to finite basis functions while Regress Now has error attributed to two sources, first, due to finite basis function and second, pathwise estimation of the conditional expectation.

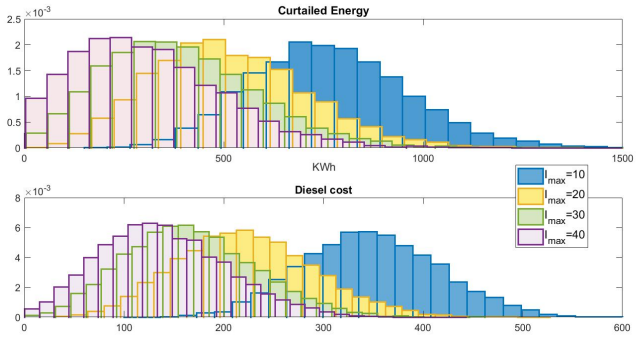


FIGURE 6. In the figure above we show histograms for different levels of battery capacity. In the top panel we display the estimated probability density of the curtailed energy, while in the bottom panel the estimated density of the cost of operating the diesel generator. Notice that the decrease in cost and curtailed energy per KWh of additional capacity is smaller for high capacity batteries.

5.2. **System behavior**

In the previous section we selected Grid Discretisation to be the best performing algorithm by our criteria. In the following we shall always employ Grid Discretisation to conduct our study of the sensitivity of the control policy and the associated cost of managing the grid to some of the parameters of the model.

The aim of the section is to build a solid understanding of the behavior of the microgrid in order to get an insight into the optimal design of the system. We decided to study the following aspects of the grid: battery capacity, represented by I_{max}; different proportion of renewable production, via the volatility σ and the mean reversion b; tenable behavior of the policy, via the switching cost K and curtailment cost C.

In order to be able to carry out our analysis, without introducing cumbersome economic and engineering details regarding the microgrid components, we have to make very simplistic assumptions. Our aim is however to guide the reader through a methodology that can be replicated to study real world microgrid systems.

5.2.1. *Battery capacity*

We study first the behaviour of the system relatively to changes in the capacity of the battery. We would expect to observe negative correlation between the quantity of diesel consumed and the battery size. We display in Figure 6 both the quantity of energy curtailed and the cost of running the diesel generator for different values of the battery capacity. We can observe that, as expected, increasing the size of the battery leads to lower diesel usage thanks to the higher proportion of renewable energy that is retained within the system. As the capacity of the battery reaches 30/40 KWh, we start observing a decrease in the cost-reduction per KWh of additional capacity suggesting that further analysis should be run in order to understand up to which size it is worth to pay to add storage capacity to the system.

We show now how to infer information about the optimal sizing of the battery, minimizing the trade off between the installation cost of a bigger battery and the reduced use of the diesel generator. Consider however that including battery ageing in the stochastic control problem is outside the scope of this paper but rather in this section we present only a post-optimization analysis. Assuming that the microgrid runs under similar conditions for the next 10 years, we can quickly estimate the total throughput of energy for the different battery capacities. Consider now that a battery does not have an infinite lifetime, but rather it should be scrapped after equivalent 4000 cycles (amount of energy for one full charge and discharge). Under the previous assumptions, we can compute how many batteries would be necessary to cover the next 10 years of operations. Similarly, using the data relative to the usage of diesel generator for different levels of capacity, we can compute the operating cost of the diesel generator over the same time period. Further exploiting the assumption about the lifetime of a battery, we obtain the cost of running the grid for 10 years as a function of the number of batteries. To conclude,

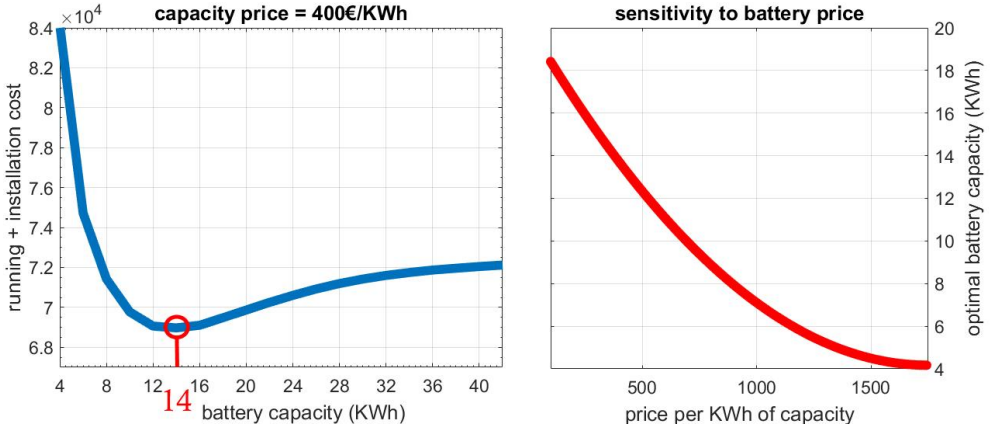


FIGURE 7. In the figure above we compute the total cost of installing and running the grid for ten years, assuming we replace the battery every 4000 cycles, and plot it against the battery capacity (left panel). From the corresponding minimum we can work out the optimal battery capacity and, further, compute the sensitivity of such result with respect to the cost per KWh of capacity.

assuming a linear cost of 400 €/KWh of capacity, we work out the installation cost of the different-size storage devices.

Once this information is collected we search for the minimum of the sum of installation and running cost and, in turn, we compute the optimal capacity. Figure 7, on the left, displays a graphical summary of the procedure just described and shows that in our problem the optimal size of the battery is 14 KWh under the current set of assumptions. Further, we study how much our result is affected by the cost per KWh of capacity, repeating the procedure above. We find that, as expected, as cost increases the size of the optimal battery decreases. Figure 7, on the right, displays such behaviour.

5.2.2. *Renewable penetration*

In this section we want to investigate how robust the microgrid is to higher penetration of renewable generation, or, in other words, to what extent the algorithm can cope with increasing randomness and decreasing predictability of the system. To model this phenomena we assume that greater penetration of renewables can be modeled by increasing both the parameters for volatility σ and the mean reversion rate λ. Increasing these two parameters makes the problem more difficult to solve, given that the control policy can rely less and less on the statistical properties of the process X which approaches white noise as high variance and high mean reversion make the current position of the process not very informative to predict its next one.

In order to establish the real added value provided by our stochastic optimization algorithm, we compare the estimated policy with an heuristic myopic control which can be reproduced in our model solving the dynamic programming equation (12) taking constant conditional expectation with respect to the control (greedy policy with respect to the current cost), particularly $\tilde{\mathcal{C}} = 0$. We plot the value of the two control policies as function of the increasing learning difficulty in Figure 8 where we observe that the importance of accounting for the future conditional expectations $\tilde{\mathcal{C}}$ increases as the predictability of X decreases.

In figure 8 we present cost of diesel (solid and dashed blue, mostly decreasing with y-axis on the left) as a function of σ for myopic and stochastic policy. The orange line (mostly increasing and y-axis on the right) represents the percentage improvement. Since increasing σ alters the volatility of the distribution of the process X, we define the mean reversion rate $\lambda := \sigma^2/(2c)$ in order to ensure that the volatility of the process is constant while we increase σ. The stochastic policy leads to at least 12% reduction in the cost of the diesel usage, compared to the myopic policy, and the difference magnifies with increasing "fluctuations" in the process. The decreasing relationship of the cost with σ signifies the importance of the battery storage system in the

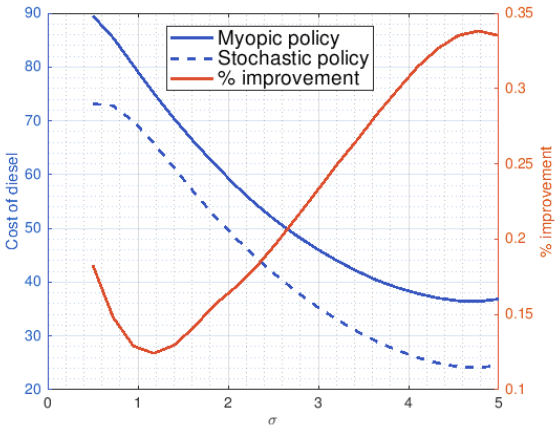
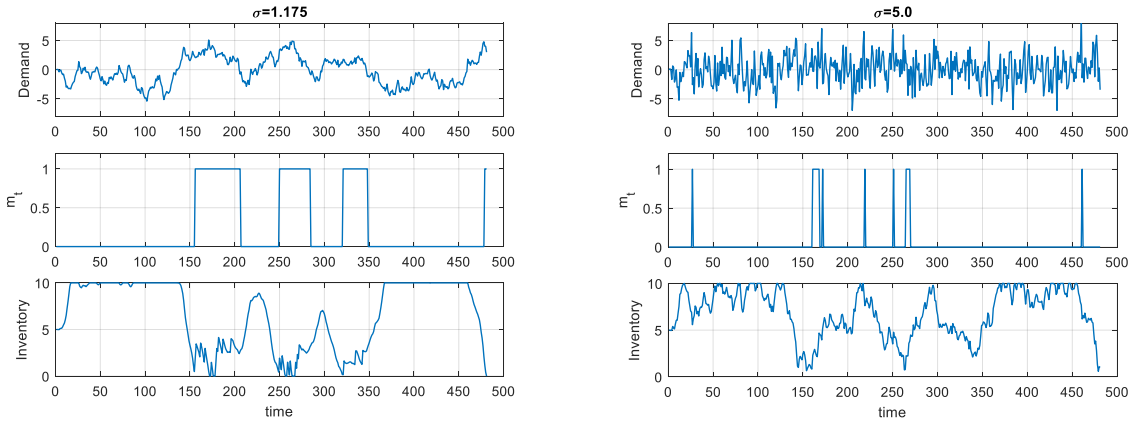


FIGURE 8. The blue lines (solid and dashed, mostly decreasing with y-axis on the left) represents the cost of the diesel usage for myopic and stochastic policy as a function of σ. The orange curve (mostly increasing with y-axis on the right) represents the percentage improvement in cost when using stochastic policy as a proportion of cost of myopic policy.



(A) (B)

FIGURE 9. Figure in the left and right panel represents demand, diesel usage and the inventory dynamics for low and high σ respectively. It is important to mention that the mean reversion rate was chosen as $\lambda := \sigma^2/8$, in order to ensure a constant volatility of the process regardless of σ. Notice the low usage of the diesel generator in the figure on the right compared to the one on the left.

microgrid which absorbs the sharp change in the demand. In figure 9 we compare the demand for two different levels of the σ, the dynamics of the diesel generator and the inventory. Notice significantly less usage of the diesel for high fluctuations, $\sigma = 5$, compared to $\sigma = 1.175$.

The results of this experiment are affected by the over-pessimistic assumption of modeling greater penetration of renewables with an increasingly unpredictable, and eventually completely random, residual demand process. This sort of analysis can however provide insight into how much (weather and load) forecasting capability will be necessary for a given level of renewable penetration.

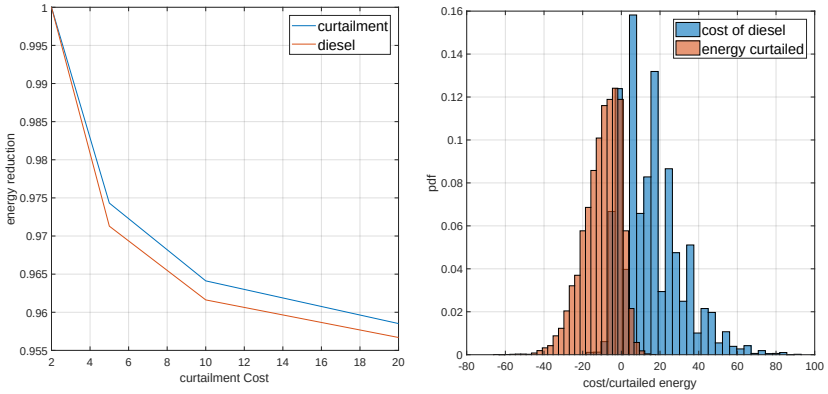


FIGURE 10. Line plot on the left, represents the impact of curtailment cost on the total curtailed energy for different C as a proportion of curtailed energy at C=2. The right panel represents the histograms for the differences between the two cases, C=20 and C=2. The blue plot represents the difference in the cost of the diesel with C=20 and C=2; the orange represents the difference in the curtailed energy. Notice the increase in curtailment cost leads to reduced curtailed energy but at the expense of inefficient diesel usage.

5.2.3. *Switching and curtailment*

We conclude this section by analyzing the dependence of the system behavior on two key parameters in the model: switching cost K and curtailment cost C. Switching cost is a system's property and the microgrid controller has little freedom over, however the controller can significantly reduce the amount of curtailed energy by choosing the appropriate curtailment cost. In figure 10, we observe that increasing the curtailment cost reduces the total curtailed energy by approximately 4%. However, it comes at the cost of inefficient usage of the diesel generator, which is represented on the right in the figure 10. The blue historgam represents the difference between the cost of diesel usage for C=20 and C=2. Similarly the orange histogram represents the difference between the energy curtailed for the two cases. Positive diesel cost depicts inefficient usage of the diesel at C=20 compared to C=2. Depending upon the specific cost functional for the diesel, the controller can use an artificial C as a parameter in the algorithm to achieve better quality of the optimization.

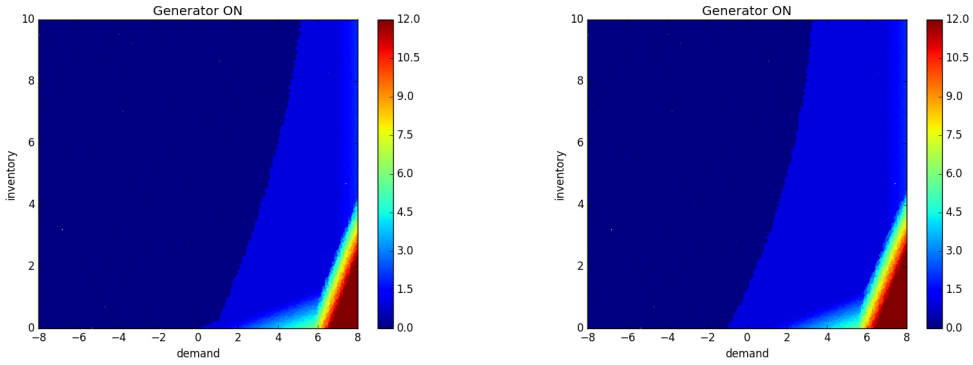
The optimal policy when the generator is ON $m_t = 1$ is significantly altered depending upon the switching cost. For example, in figure 11, we present the control maps associated with K=2 and K=5. As expected, larger switching cost disincentivise the controller to switch OFF the diesel generator once it's ON. However, we don't observe "significant" change in the control policy due to increase in switching cost when the generator is OFF.

6. Comparison with deterministically trained policy

In this section we compare our stochastic optimization algorithm with a deterministically trained policy. The latter is widely used in online optimization where the solution is computed with respect to the best forecast available at a given time. We emulate this situation by computing the optimal set of actions for a particular deterministic demand trajectory at different levels of the inventory. We assume that the forecast of the demand is given by:

$$X_{t+1} = X_t + 0.5(6\sin(\frac{\pi t}{12}) - X_t)\Delta t; \quad t \in \{0, 1, \ldots, T-1\}. \tag{17}$$

Equation (17) implies periodicity of one day in the residual demand and is equivalent to $\sigma = 0$, $b = 0.5$ and $\Lambda_t = 6\sin(\frac{\pi t}{12})1$ in (2). Zero volatility in the residual demand curve leads to a deterministic optimal control problem, rather than a stochastic control problem we have presented in section 5.



(A) $K = 2$ (B) $K = 5$

FIGURE 11. Figure on the left represents the control map for switching cost $K = 2$, while the figure on the right represents the control map for $K = 5$ when the generator is ON. Notice the increase in area for light blue (corresponding to $d = 1$) in the figure on the right because of increased switching cost.

Notice that the deterministic optimal control problem results in a sequence of control maps $d_t : (w, m) \to [d_{min}, d_{max}] \cup 0$. As a result, although the policy has been trained on a deterministic residual demand, it dynamically adapts itself to different inventory levels and state of the diesel generator, when tested in a stochastic environment. We present the modified algorithm in 4. There are two key differences from the previous algorithm, first, we use one dimensional projection of the value function and second, we replace regression with interpolation since there is no randomness left in the problem.

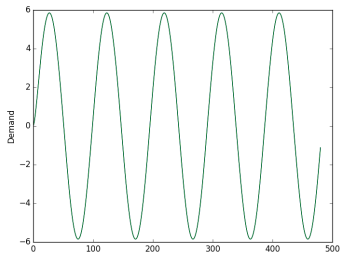
Algorithm 4 Regression Monte Carlo algorithm for deterministic demand

Simulate $\{X_t\}_{t=1}^N$ according to its dynamics;
Discretize I_t into M levels indexed by j s.t. $\{I_t^j\}_{j=1}^M$;
Initialize the value function $V(T, I_T^j, m_T) = g(I_T^j), \quad \forall j = 1, \ldots, M$ and $m_T = \{0, 1\}$;
for $t = N - 1$ to 1 **do**
    Find interpolation function $\mathcal{B}(t + 1, I_{t+1}, m)$ for $\{V(t + 1, I_{t+1}^j, m_{t+1})\}_{j=1}^M \quad$ for each $m = 0, 1$
    Compute the set of admissible controls as $\mathcal{U}_t$
    for $j = 1$ to M **do**
        for $m = 0$ to 1 **do**
            $F = \mathcal{B}(t + 1, I_t^j, 0)$

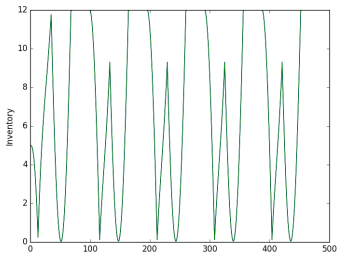
$$V(t, I_t^j, m) = \begin{cases} \min\limits_{d \in \mathcal{U}_t \setminus \{0\}} \left\{ p\rho(d) + CS_t \mathbb{1}_{\{S_t < 0\}} + \mathcal{B}(t + 1, I_t^j - B_t^d, 1) \right\} + K\mathbb{1}_{\{m=0\}} \wedge F & \text{if } 0 \in \mathcal{U}_t \\ \min\limits_{d \in \mathcal{U}_t} \left\{ p\rho(d) + CS_t \mathbb{1}_{\{S_t < 0\}} + \mathcal{B}(t + 1, I_t^j - B_t^d, 1) \right\} + K\mathbb{1}_{\{m=0\}} & \text{otherwise} \end{cases}$$

output: control policy $\{\mathcal{B}(t, \cdot, \cdot)\}_{t=2}^N$.

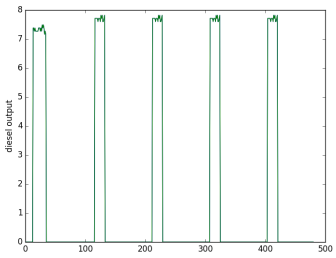
In order to understand the solution of the deterministic problem, in figure 13 we present the dynamics of the optimal control and inventory corresponding to the demand faced in (A). As expected, diesel switches on when the demand is high and it keeps it running just long enough that the battery is empty before it faces negative



(A) Demand



(B) Inventory



(C) Diesel Output

FIGURE 12. The image illustrates the dynamics of the inventory and control for the deterministic control problem. Figure (A) represents the demand in equation (17), the optimal control of the diesel in figure (C) and the corresponding dynamics of the inventory in figure (B).

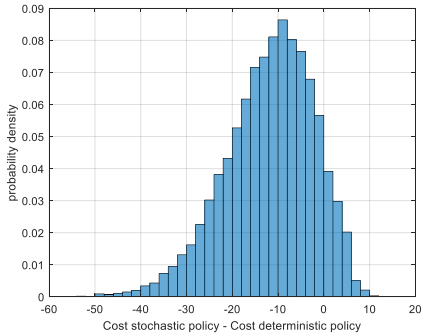


FIGURE 13. Difference of the Cost of Stochastic and deterministic policy for K=5

Switching Cost	K=2	K=5	K=10
Deterministic	138.56	162.63	201.52
Stochastic	131.86	150.49	178.22
% difference	4.84%	7.46%	11.56%

TABLE 1. Comparison of deterministic and stochastic trained policy.

residual demand to charge the battery. Moreover, there is substantial curtailment of energy since the battery is not large enough to store all the excess energy.

In order to quantify the gain due to formulating the microgrid management problem as a stochastic control rather than traditional deterministic control, we compare the performance of the deterministically trained strategy of this section to its stochastic counterpart developed in this paper. While the deterministic control problem was solved using the residual demand curve (17), the stochastic control problem was fed in with the residual demand curve (18). Finally, we test both the strategies on fresh out-of-sample paths following the residual demand (18).

$$X_{t+1} = \left(X_t + 0.5(6\sin(\frac{\pi t}{12}) - X_t)\Delta t + 2\sqrt{\Delta t}\xi_t\right) \wedge 10\,; \quad t \in \{0, 1, \ldots, T-1\} \tag{18}$$

In figure 13, we present the histogram of the cost from the stochastic policy and the deterministic policy pathwise for 10,000 out-of-sample paths. As evident, most of the distribution lies on the negative side, implying gain due to stochastic policy. To measure this difference, in table 1, we quantify the gain of the stochastic policy for different switching cost. For switching cost of K=5, we observe that the stochastic policy is 7.5% better than the deterministic policy. As the switching cost increases, mistakes made by deterministic policy become more expensive leading to higher percentage difference.

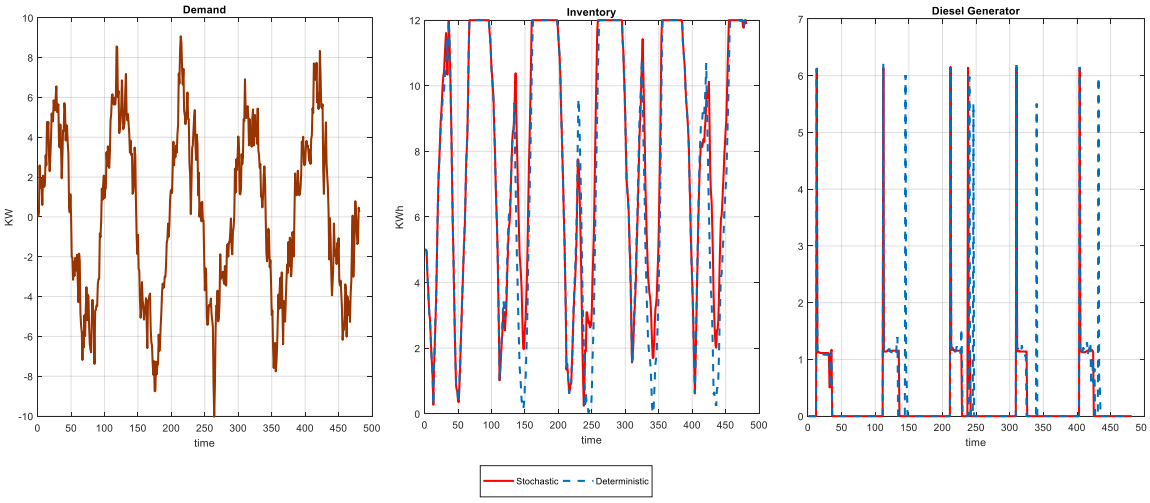


FIGURE 14. The figure above presents the pathwise comparison of stochastic and deterministic policy for the same demand on the left panel. The center panel represents dynamics of the inventory due to control on the right panel. Particularly notice the difference in switching times for the diesel in the deterministic policy and stochastic policy.

Finally, Figure 14 displays the behavior of inventory and the cost along a random trajectory of residual demand. In blue we show the stochastically trained control policy and in orange the deterministically trained. The stochastic policy has lesser switch of the diesel generator and thus lower costs. The spikes in the cost function for the deterministic policy is due to poor management of the inventory and thus inefficient usage of the microgrid.

7. Conclusion

In this paper we solved the problem of optimal management of a microgrid by employing three algorithms from the Regression Monte Carlo literature, namely: Regress Now, Regress Later and Inventory Discretization. We find that Inventory Discretization significantly outperforms the other two methods. Besides algorithm design, we propose a methodology to optimize the design of the grid and determine the optimal sizing of the battery. In addition, we perform a thorough sensitivity analysis to some of the key parameters, showing the robustness of our solution. Finally, we compare the control policy estimated by our algorithm to industry standard deterministic control, observing a 5-10% reduction in cost.

Future research in this direction will include further studies of the optimal sizing of the battery by explicitly incorporating the wearing off caused by usage. Another more challenging direction is to understand the impact of delay, e.g., in the switching of the diesel generator, on the optimal management of the microgrid. This problem introduces several mathematical and algorithmic issues which are currently the focus of our research.

8. Acknowledgements

This research was supported by the FIME Research Initiative. The research of C. Alasseur and X. Warin has also benefited from support by the ANR project CAESARS (ANR-15-CE05- 0024). The research of Peter Tankov has also benefited from support by the ANR project FOREWER (ANR-14-CE05- 0028). The research of Aditya Maheshwari has benefited from support by NSF DMS-1736439.

References

[1] Alessandro Balata and Jan Palczewski. Regress-Later Monte Carlo for optimal control of Markov processes. *ArXiv e-prints*, December 2017.

[2] Alessandro Balata and Jan Palczewski. Regress-Later Monte Carlo for Optimal Inventory Control with applications in energy. *ArXiv e-prints*, March 2017.

[3] Dimitri P. Bertsekas. *Stochastic Optimal Control*. Academic Press, New York, 1978.

[4] Alexander Boogert and Cyriel de Jong. Gas storage valuation using a monte carlo method. *The Journal of Derivatives*, pages 81–98, March 2008.

[5] Bruno Bouchard and Xavier Warin. Monte-carlo valuation of american options: facts and new algorithms to improve existing methods. In *Numerical methods in finance*, pages 215–255. Springer, 2012.

[6] René Carmona and Michael Ludkovski. Valuation of energy storage: an optimal switching approach. *Quantitative Finance*, 10(4):359–374, 2010.

[7] Jérôme Collet, Olivier Féron, and Peter Tankov. Optimal management of a wind power plant with storage capacity. HAL preprint hal-01627593, 2017.

[8] Huajie Ding, Zechun Hu, and Yonghua Song. Stochastic optimization of the daily operation of wind farm and pumped-hydro-storage plant. *Renewable Energy*, 48:571–578, 2012.

[9] Huajie Ding, Zechun Hu, and Yonghua Song. Rolling optimization of wind farm and energy storage system in electricity markets. *IEEE Transactions on Power Systems*, 30(5):2676–2684, 2015.

[10] Hugo Gevret, Nicolas Langrené, Jerome Lelong, Xavier Warin, and Aditya Maheshwari. STochastic OPTimization library in C++. Research report, EDF Lab, May 2018.

[11] Pierre Haessig, Bernard Multon, Hamid Ben Ahmed, Stéphane Lascaud, and Pascal Bondon. Energy storage sizing for wind power: impact of the autocorrelation of day-ahead forecast errors. *Wind Energy*, 18(1):43–57, 2015.

[12] Naoki Hayashi, Masaaki Nagahara, and Yutaka Yamamoto. Robust ac voltage regulation of microgrids in islanded mode with sinusoidal internal model. *SICE Journal of Control, Measurement, and System Integration*, 10(2):62–69, 2017.

[13] B. Heymann, J. F. Bonnans, F. Silva, and G. Jimenez. A stochastic continuous time model for microgrid energy management. In *2016 European Control Conference (ECC)*, pages 2084–2089, June 2016.

[14] Benjamin Heymann, J. Frédéric Bonnans, Pierre Martinon, Francisco J. Silva, Fernando Lanas, and Guillermo Jiménez-Estévez. Continuous optimal control approaches to microgrid energy management. *Energy Systems*, Jan 2017.

[15] Hao Liang and Weihua Zhuang. Stochastic modeling and optimization in a microgrid: A survey. *Energies*, 7(4):2027–2050, 2014.

[16] Michael Ludkovski and Aditya Maheshwari. Simulation Methods for Stochastic Storage Problems: A Statistical Learning Perspective. *ArXiv e-prints*, 2018.

[17] S. Mashayekh, M. Stadler, G. Cardoso, M. Heleno, S. Chalil Madathil, H. Nagarajan, R. Bent, M. Mueller-Stoffels, X. Lu, and J. Wang. Security-constrained design of isolated multi-energy microgrids. *IEEE Transactions on Power Systems*, PP(99):1–1, 2017.

[18] Salman Mashayekh, Michael Stadler, Gonçalo Cardoso, and Miguel Heleno. A mixed integer linear programming approach for optimal der portfolio, sizing, and placement in multi-energy microgrids. *Applied Energy*, 187(Supplement C):154 – 168, 2017.

[19] Lanre Olatomiwa, Saad Mekhilef, A.S.N. Huda, and Olayinka S. Ohunakin. Economic evaluation of hybrid energy systems for rural electrification in six geo-political zones of nigeria. *Renewable Energy*, 83(Supplement C):435 – 446, 2015.

[20] Daniel E Olivares, Ali Mehrizi-Sani, Amir H Etemadi, Claudio A Cañizares, Reza Iravani, Mehrdad Kazerani, Amir H Hajimiragha, Oriol Gomis-Bellmunt, Maryam Saeedifard, Rodrigo Palma-Behnke, et al. Trends in microgrid control. *IEEE Transactions on smart grid*, 5(4):1905–1919, 2014.

[21] S. Surender Reddy, Vuddanti Sandeep, and Chan-Mook Jung. Review of stochastic optimization methods for smart grid. *Frontiers in Energy*, 11(2):197–209, Jun 2017.

[22] Xavier Warin. *Gas Storage Hedging*, pages 421–445. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.