This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

Adaptive Trading in Continuous Intraday Electricity Markets for a Storage Unit

Gilles Bertrand, *Student, IEEE* and Anthony Papavasiliou, *Member, IEEE*

***Abstract*—The increasing integration of renewable resources in electricity markets has increased the need for producers to correct their trading position close to real time in order to avoid volatile real-time prices. The closest option to delivery time in European markets is to trade in the continuous intraday market. This market is therefore an attractive trading outlet for assets that target at extracting value from their flexibility. Trading in this market is challenging due to the multistage nature of the problem, its high uncertainty and the fact that decisions need to be reached rapidly, in order to lock in profitable trades. We model the trading problem of a storage unit in the Markov Decision Process framework. We present an approach based on policy function approximation for tackling the problem. We provide relevant parameters for defining our policy, and demonstrate the effectiveness of our approach by comparing it to the rolling intrinsic policy on real historical data. Our proposed approach outperforms the rolling intrinsic policy, which is commonly employed in practice for storage units, by increasing profitability by 17.8% on out-of-sample testing for a storage with perfect round-trip efficiency and by 13.6% for a storage unit with a round-trip efficiency of 81%.**

***Index Terms*—Markov decision processes, policy function approximation, reinforcement learning, continuous intraday market.**

I. Introduction

Following the introduction of the climate and energy package in Europe [1], the integration of renewable energy in Germany has increased from 18.2% in 2010 to 32.2% in 2016 [2]. These renewable resources increase the variability of supply in the market, and consequently increase the need for correcting system dispatch closer to real time. An interesting option for such corrections is to trade in the continuous intraday market (CIM), which explains the recent increase of liquidity in this market. Specifically, traded volumes in the German CIM have increased from 10 TWh in 2010 to 45 TWh in 2016 [3]. This market is therefore becoming an interesting option for fast-moving assets, such as batteries or pumped hydro storage, to extract value from their flexibility.

Several papers analyze the optimization of bidding strategies in different electricity markets. In [4], the authors consider trading in the day-ahead market and covering their position in imbalance for a wind power producer. This work has been extended in [5] in which the authors also consider bidding in the intraday market. In [6], the authors develops a trading strategy for a wind power producer who trades in the day-ahead market, followed by settlement in the real-time market. The authors account for the impact of the dependence between the wind production error and the real-time price on the trading strategy of the wind farm.

A certain body of the literature focuses specifically on storage units. The operation of storage units in the context of a US-style centralized unit commitment has been studied in the literature using unit commitment models such as in [7] and [8]. Nevertheless, these models are out of scope in an EU context, where resource owners self-commit and self-schedule individual resources at the nomination stage which follows the clearing of the portfolio-based day-ahead market. In the EU context, the authors in [9] focus on the interaction of trading strategies in the day-ahead market and the balancing market, while the interaction between day-ahead and intraday auctions has been analysed in [10].

The strategies developed for these markets cannot be applied directly to the CIM due to the continuous format of this market, which differs from the day-ahead auction or the intraday auction. Indeed, in auctions the producer has one chance to submit bids. Instead, in the CIM, the producer is afforded a certain amount of time in order to observe the offers submitted by other participants. Moreover, in the CIM, buy and sell prices for the same delivery time may evolve over the horizon of trading. Due to these particularities, the CIM has received separate treatment in the literature.

The specific literature about the CIM can be classified into the three following categories.

(i) The first category of papers focuses on modeling the price evolution in the CIM. This includes literature that focuses on the explanatory variables for the evolution of the price [11], [12], and on the factors that influence the liquidity and the bid-ask spread [13]. In [14] the authors develop a Hawkes process for modeling the arrival of orders. A model for the simulation of the CIM based on data from the European Power Exchange is proposed in [15].

(ii) The second category of papers focuses on optimal trading strategies, and assumes that the intraday prices follow a given parametric model. Trading for a pumped hydro storage facility is presented in [16] and [17]. The first paper discusses the optimization problem of pumped hydro storage trading, where it is assumed that traders can access a forward curve. The second paper studies the problem of trading in the CIM and in the balancing market. Other papers develop trading strategies for other types of asset. In [18], the authors consider trading in the CIM for balancing the forecast error of renewable energy. The authors assume that the intraday price follows a geometric Brownian motion. A trading strategy for a thermal power plant is developed in [19], where it is assumed

Gilles Bertrand is with CORE, UCLouvain, Louvain la Neuve, Belgium gilles.bertrand@uclouvain.be

Anthony Papavasiliou is with CORE, UCLouvain, Louvain la Neuve, Belgium anthony.papavasiliou@uclouvain.be

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

that the intraday price follows an additive Brownian motion. This price is further influenced by the most recent trades of the producer.

(iii) The third strand of literature focuses on developing trading strategies, without placing assumptions on the parametric distribution of the data. In [20] the authors propose a heuristic trading method for wind power producers. The authors in [21] consider the problem of trading without assets in the CIM in order to cover a position in the imbalance market. The authors model the problem as a one-stage MDP, and solve it using policy functions. Related to MDP, two papers have modelled the problem of trading for a storage unit in the CIM using MDP [22], [23]. The first one relies on value function approximation. The second one resorts to policy function approximation in the form of a threshold policy in order to simplify the problem.

In the present paper, we consider a generalization of the problem that is presented in [23], where we additionally account for round-trip efficiency losses of storage units. The contributions of this paper are the following: (i) We cast the intraday market trading problem for a storage unit in the MDP framework. (ii) We employ policy function approximation in order to arrive at a computationally tractable problem formulation. More precisely, we use a threshold policy according to which we seek a sell threshold above which we accept to sell power, and a buy threshold below which we accept to buy power. (iii) We propose a parametrization of the trading thresholds that accounts for several effects, in order to arrive at a policy that outperforms a benchmark policy referred to as *rolling intrinsic*. (iv) We analyse the results at higher trading frequency than the one considered in [23]: whereas in [23] the results are derived using hourly frequency, in the present publication we consider a frequency of 5 minutes for learning and 1 second for testing out of sample. Moreover, we demonstrate through experiments the important role of frequency on the training and evaluation of trading strategies.

Section II describes the operation of the continuous intraday market and how we simulate it. Section III explains how to model the trading problem faced by a storage unit in the MDP framework. In section IV, we introduce the idea of a threshold policy, in order to arrive to a tractable problem for optimizing over policies. We also recall the REINFORCE algorithm for optimizing the policy function parameters. Section V presents the factors that we propose in order to adapt the threshold policy. In section VI, we present a test case which demonstrates the effectiveness of our approach on German market data, and we analyze how our proposed policy fares relative to rolling intrinsic. Finally, in section VII we conclude the paper and propose directions for further research.

II. Continuous intraday electricity markets operation

In this section, we describe the operation of a continuous intraday market. We base our description on the German market, which is representative of the operation of electricity markets in Central Europe.

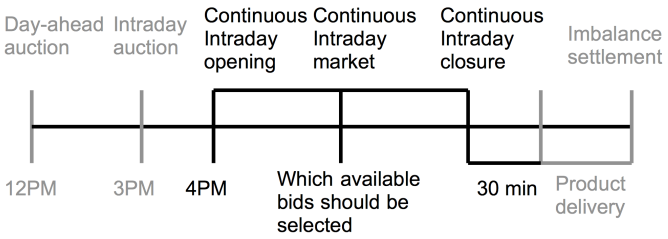


Fig. 1: The sequence of operations in a typical central European short-term electricity market

A. Short-term electricity market description

In Fig. 1 we describe the timing of the different short-term electricity markets in a typical central European power exchange. We use indicative values for the German market. Short-term market trading commences with the day-ahead market at 12 noon, the day before electricity delivery (D-1) [24]. Subsequently, the intraday *auction* takes place at 3PM on D-1 [25]. Following the conclusion of the intraday auction, the continuous intraday market (which is a *separate* process from the intraday auction) commences at 3 PM for hourly products and at 4 PM for quarter-hourly products [26]. The CIM closes 30 minutes before delivery. Finally, the imbalance is cleared at the imbalance price [27].

In this paper, we consider trading only in the continuous intraday market. In the CIM, at any moment, the traders have access to an order book. This order book is the list of all available bids on the CIM. At any moment, a trader can submit new bids or accept (partially or fully) existing bids. The bids available on the market have 4 characteristics: (i) a delivery time, which is the moment at which the power should be injected to or withdrawn from the grid; (ii) the type of bid (sell/buy): a sell (resp. buy) bid corresponds to an offer from a counter-party to sell (resp. buy) power; (iii) the price of a bid (in €/MWh); and (iv) the quantity of a bid (in MWh).

We are interested in the development of trading strategies for a storage asset owner. A storage unit is an especially interesting asset to consider in the context of intraday trading, since it offers the possibility to procure power from relatively cheap sell bids, store the power, and sell it back to subsequent buy bids that place a greater valuation on the procured power. We consider the following simplifications in the sequel:

1) The trading strategies that we develop are balanced. This implies that, at the closing time of the continuous intraday market, the position of the storage unit should be feasible. We thus adhere to German regulation [27], which requires that the producer can only be in imbalance if this imbalance is caused by an unpredictable event[1] (forecast error, outages). Practically, this implies that if we do not have any energy stored in our reservoir, we cannot sell power and cover it in the imbalance market.

[1]Note that US market operations differ in this respect due to central dispatch, which allows the system operator to actively manage resources in real time in order to increase trading surplus, as opposed to requiring them to remain in balance at all costs.

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

2) We only accept bids that are already present in the market, as opposed to also placing bids in the market. Adding the option of placing bids would complexify our Markov Decision Process in 2 ways: (i) We would have to add to our state all the bids that we have placed on the market at previous time steps. (ii) We would need to extend our action space in order to decide on suppressing the bids that we have placed at previous time steps.
3) In practice, CIM bids are categorized into more complex products, referred to as continuous bids, all-or-none bids, block bids, iceberg bids, and so on [26]. For our case study, we assume that all the data that we have access to corresponds to continuous bids. This implies that we can accept fractions of bids. There are two reasons for adopting this simplification: (i) The information about the type of bids (continuous, integer, block) is not disclosed in the German market data set that we use for our case study. (ii) Practitioners have indicated to us that the impact of this restriction is minor, because most of the bids are continuous bids. To a certain extent, the more complex products have been inherited from the products that are available in the day-ahead market. A major reason for the existence of these complex products in the day-ahead market is in order to provide the option for a producer to account for complex unit commitment constraints. This interest is more limited in the CIM, because the commitment variables have to be decided several hours before delivery, through the so-called nomination procedure.
4) We only consider hourly products in our paper, as opposed to also considering quarterly products that refer to delivery within a specific 15-minute interval.
5) We assume that our producer is risk-neutral. The reason for this is that the daily average profit obtained for our storage unit is around 6400 €, whereas the profit for the worst day is approximately -500 €. Typical energy companies have the financial ability to absorb this potential loss for several days without any problem. Therefore, the company can only focus on maximizing its long-term profit, which will be obtained by being risk neutral on a daily basis.
6) We assume that, no matter which bid we accept in the market, we do not influence the bids that the other actors will place later in the market. This simplification has been adopted in order to simplify the problem, and is completely in line with the state of the art on the topic of intraday trading in electricity markets [16], [20]. Moreover, we have assessed the validity of this assumption in the electronic supplement[2].

B. Market simulation

In order to simulate the evolution of the order book, we consider 4 types of events:

1) **Open**: the appearance of a trade
2) **Cancel**: the disappearance of a trade
3) **Acceptance**: the acceptance of a certain quantity of a bid
4) **Trading**: the moment when we decide which bids we accept.

The simulation of the market can now be described as follows. At the beginning of the simulation, we rank all the events, which are included in the set **Event**, chronologically. We then iterate on this set: for each new event j, we classify it in one of the 4 categories and we update the order book as described in the following procedure.

$L = []$
for $j \in$ Event
    if $j \in$ **Open**
        Add bid j to L
    elseif $j \in$ **Close**
        Remove bid j from L
    elseif $j \in$ **Acceptance**
        Reduce partially accepted quantity from bid j
    elseif $j \in$ **Trading**
        Launch the trading algorithm
        Remove the bids that we have accepted from L
    end
end

III. Modelling the intraday trading problem using the MDP framework

Having defined how to simulate the market, we can now analyse the trading problem. The decision problem is to decide, at different moments of the Continuous Intraday Market, which bids should be accepted in order to maximize the future expected profit of our storage unit. In the rest of the paper, we refer to a general storage unit. This storage unit is characterized by a certain charging and discharging efficiency. These settings create the basis for representing a battery, a simplified model of a pumped storage hydro unit, or certain types of demand response. The main trade-off for our decision problem is the following: *Do we want to trade power at the current price and lock in the profit? Or is it worth waiting for a potential future bid the price of which would be more advantageous, despite the risk that the current favorable bids may disappear?* Our decision problem falls under the scope of multistage optimization under uncertainty, because we need to arrive to decisions knowing that recourse actions can be adopted in an uncertain future. A common way to approach this class of problems is by using the Markov Decision Process framework. In order to characterize an MDP, we need to model the state variables, the action variables, the reward and the state transition function.

A. State variables

In order to reach a decision at time step t, we require 3 ingredients in our state S_t: (i) The offers available in the continuous intraday market at time step t. This data is available

[2]The electronic supplement is available at the following link: https://sites.google.com/site/gillesbertrandresearch/publications/app-transaction-2019

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

in the market order book. (ii) A variable $v_{t-1,d}, \forall d \in D$ which indicates the capacity that would be stored in the storage unit at delivery hour d if we were only executing the trades decided at time step $t-1$ or earlier. This value can be easily computed based on the results of all the trades that we have realized in the past. (iii) Exogenous data that we anticipate should influence our decision. Some examples of these exogenous parameters include the remaining time before market closure, and the price of the intraday auction. The full list of these parameters, and the way in which we use them, is discussed in Section V.

B. Action variables

In order to model our action A_t, we require one action variable $a_{t,d}$ for each delivery time d. This action indicates how much we wish to sell at time step t. In theory, this variable can be continuous. But, in order to reduce the size of the action space, we will discretize this variable into $2n+1$ potential actions:

$$a_{t,d} \in \{-q_n, \cdots, -q_1, 0, q_1, \cdots q_n\}$$

C. Reward

The total reward obtained from the CIM at time step t is equal to the sum of the rewards obtained for every delivery hour:

$$R_t(S_t, A_t) = \sum_{d \in D} \text{rev}(a_{t,d}),$$

where the reward for delivery hour d at time step t is computed as the integral of the demand curve $p_{t,d}$ from 0 to $a_{t,d}$:

$$\text{rev}(a_{t,d}) = \int_0^{a_{t,d}} p_{t,d}(z)dz.$$

D. State transition function

In the case of the intraday trading problem, we assume that there exists a state transition function but that it is unknown (since we do not place any assumptions on the evolution of intraday prices). This prohibits us from using methods such as policy iteration or value iteration. Nevertheless, Reinforcement Learning techniques are perfectly suitable for such a setting. Indeed, the idea of Reinforcement Learning techniques is to gain knowledge about the environment by running episodes of the task (in our case, each episode corresponds to a day of trading). Note that the round-trip efficiency of a storage unit is part of this transition function, which we do not model explicitly.

IV. Policy function approximation

We are interested in an optimal policy for trading. A policy is a function which is a distribution over actions for every state of the MDP. The policy should be selected among a set of policies Π, such that we maximize the future expected reward in Eq. (1)

$$\max_{\pi \in \Pi} \sum_{t=1}^{T} \mathbb{E}\left[R_t(S_t, A^\pi(S_t))\right], \tag{1}$$

where $A^\pi(S_t)$ is the action taken if we are in state S_t and we follow the policy π.

In our case, we have infinite states, since the prices for the different delivery times are continuous. Therefore, the problem of finding an optimal policy becomes infinite-dimensional [28], [29]. Thus, the problem as expressed in its initial formulation is intractable.

In order to obtain an approximate solution to the problem, we resort to policy function approximation. The idea of policy function approximation is to express the policy $\pi_\theta(a|s)$ with respect to a parameter vector θ, and to optimize over this θ:

$$\pi_\theta(a|s) = \mathbb{P}[A_t = a | S_t = s; \theta].$$

We thus restrict the policy domain, which implies that we will obtain a policy which may be sub-optimal, which is the cost of restricting our search over θ. The remainder of this section explains how we calibrate the weights θ on the basis of repeated episodes of trading and how we implement a threshold policy for our trading problem.

A. REINFORCE algorithm

In order to optimize the parameter vector θ, we use the REINFORCE algorithm:

- Initialize θ
- for each episode $\{s_1, a_1, r_2, \cdots, s_{T-1}, a_{T-1}, r_T\} \sim \pi_\theta(a|s)$
  for t = 1 : T-1 do
  $$\theta = \theta + \gamma_t \nabla_\theta \log(\pi_\theta(a|s)) g_t \tag{2}$$
  end
  end

The REINFORCE algorithm adapts the parameter vector θ so as to maximize expected rewards from a certain policy, based on repeated episodes of the decision process. An episode corresponds to one day of trading. The episode commences at the first trading interval of the day. Given a state s_1, we select an action based on our policy function, we collect a reward r_2, and we arrive at the state s_2. This process is repeated until the end of the trading day. When the episode is finished, we update θ using Eq. (2), where g_t is the profit from t until the end of the episode T. It has been proven in [30] that the REINFORCE algorithm is effectively a stochastic gradient algorithm. It is therefore guaranteed to converge under standard stochastic approximation conditions for decreasing step-sizes γ_t.

B. Threshold policy

We focus on a policy which is parametrized by buy and sell price thresholds. The threshold policy that we investigate in this paper accepts sell bids if their price is below a buy threshold, and accepts buy bids if their price is above a sell threshold. Our focus on threshold policies is justified

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

by several factors: (i) Optimal inter-temporal arbitrage in a deterministic setting is achieved by a threshold policy, as proved in [23]. This result has been extended to a three-stage stochastic program in the electronic supplement of this paper. (ii) Threshold policies have also been *proven* to be optimal in a number of papers in the literature regarding specific instances of stochastic optimal control problems with uncertain prices [31]–[33]. (iii) The idea of using a threshold policy in order to trade for a storage unit has already been proposed in other settings [34].

We apply a stochastic threshold policy, in order to ensure sufficient exploration to take place during the learning stage of the algorithm. Concretely, we propose drawing the sell and buy thresholds from a Gaussian distribution. Therefore, we define our policy parameter, θ, as $\theta = (\mu_X, \sigma_X, \mu_Y, \sigma_Y)$,where the buy threshold for delivery hour d, X_d, is drawn according to a normal distribution with parameters[3] $(\mu_X, \exp(\sigma_X))$, and the sell threshold for delivery hour d, Y_d, is drawn according to a normal distribution with parameters $(\mu_Y, \exp(\sigma_Y))$. We draw one independent threshold per delivery hour d, therefore an action at delivery time d_1 is independent of the action at delivery time d_2. By independence, the distribution of actions over all future delivery hours can be expresses as:

$$\pi_\theta(a|s) = \prod_{d \in D} \pi_\theta^d(a|s)$$

In order to illustrate how the stochastic threshold is implemented, we consider the example of Fig. 2. at delivery hour d. (i) The solid black decreasing function (solid line) corresponds to the buy bids that are available in the order book for delivery hour d. This data is available in the order book at the time we are deciding on whether or not to accept a bid. The demand curve is associated with the lower x-axis. (ii) The bell curve represents the probability density function of the threshold. This curve can be computed based on the current vector parameter θ. The bell curve is associated with the upper x-axis. With these two elements, we illustrate how we use the threshold policy in order to arrive at decisions. Consider, for instance, the action *Sell* 10 MWh: if the sell threshold that we draw is between the price associated to a sell quantity of 15 MWh and the price associated to a sell quantity of 5 MWh, we sell 10 MWh. The probability of this action corresponds to the light grey surface $\pi_\theta^d(10|s)$. This probability can also be computed mathematically, as illustrated below:

$$\begin{aligned}\pi_\theta^d(10|s) &\triangleq \Pr(a_{t,d} = 10) \\ &= \Pr(p(15) \leq Y_d \leq p(5)) \\ &= \Pr(Y_d \leq p(5)) - \Pr(Y_d \leq p(15)) \\ &= \Phi(p(5); \mu_Y, \exp(\sigma_Y)) - \Phi(p(15); \mu_Y, \exp(\sigma_Y))\end{aligned}$$

where $\Phi(\cdot; \mu, \sigma)$ indicates the cumulative distribution function of the normal distribution with mean μ and standard deviation σ. In order to apply the REINFORCE[4] algorithm, we also need to compute the policy derivatives for the different actions. These derivatives can be computed analytically as illustrated below for the derivative of the probability of the action *Sell* 10 MWh with respect to μ_Y:

$$\begin{aligned}\frac{\partial \pi_\theta^d(10|s)}{\partial \mu_Y} &= \frac{\partial \Phi(p(5); \mu_Y, \exp(\sigma_Y))}{\partial \mu_Y} \\ &\quad - \frac{\partial \Phi(p(15); \mu_Y, \exp(\sigma_Y))}{\partial \mu_Y} \\ &= -\phi(p(5); \mu_Y, \exp(\sigma_Y)) \\ &\quad + \phi(p(15); \mu_Y, \exp(\sigma_Y))\end{aligned}$$

where $\phi(\cdot; \mu, \sigma)$ denotes the probability density function of the normal distribution with mean μ and standard deviation σ.

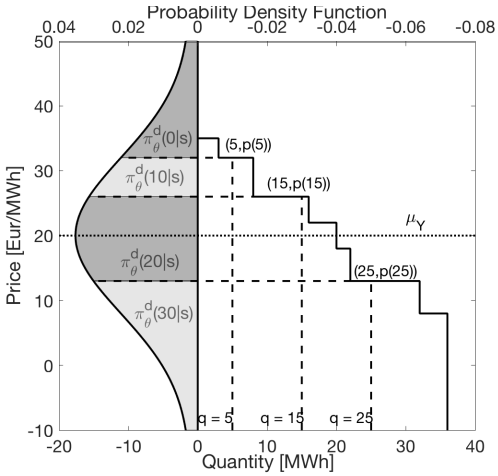


Fig. 2: Threshold policy for the hydro problem if we consider four possible actions: sell 0, 10, 20 or 30 MWh. The bell curve indicates the probability density function of the sell threshold. The two light grey segments and the two dark grey segments of the bell curve indicate the probability of each of the four actions. The solid black decreasing function corresponds to the buy bids that are available in the order book.

V. Factors driving the optimal threshold

In the previous section, we have developed a basic threshold policy for trading in the CIM. This simple threshold policy does not achieve satisfactory performance in practice, because it is not sufficient to maintain the same threshold for every time step of every day. This suggests that the threshold should be further dependent on certain factors that are pertinent towards an adaptive trading strategy. In this section, we propose a number of such factors and explain the reason for which we consider them. Then, we explain how the REINFORCE algorithm can be adapted in order to incorporate these factors.

A. Delivery time

The need for using different thresholds depending of the delivery hour is illustrated in Fig. 3. This graph represents the *CIM price* (which we define as the center of the bid-ask spread) for the 24 different delivery hours. The cross represents

[3]We use $\exp(\sigma_X)$ and not σ_X directly in order to ensure that the standard deviation remains positive.

[4]We present an example of all the computations needed in order to realize one iteration of the REINFORCE algorithm in the electronic supplement.

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

the price of buying energy at the 6^{th} hour, while the dot is the price of buying energy at the 17^{th} hour. These two prices are equal, however the buying decision should be different. Indeed, the price corresponding to the cross is not interesting, because the same amount of power could have been procured and stored at the reservoir at a lower price at hour 4. On the contrary, the price corresponding to the dot is interesting, because it corresponds to a local minimum price. Thus, in hour 17 we can buy power, in order to sell that power back at a later delivery time.

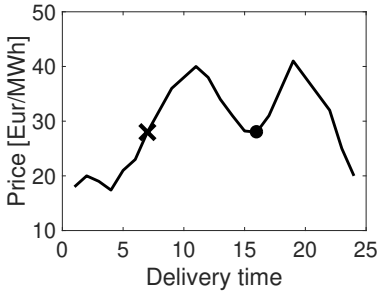


Fig. 3: The delivery time of an order impacts its threshold: buying power at 30 €/MWh is not worthwhile in hour 6, but it is worthwhile in hour 17.

Having argued that it is necessary to employ different thresholds for different delivery times, our idea is to define regimes for which the threshold mean should be the same. We will define these regimes based on the intraday auction price curve, which conveys a significant amount of information about the CIM price.

We present an example of these regimes based on the intraday auction price curve, for one precise day of our dataset, in Fig. 4. These graphs illustrate that the buy threshold switches at the maximum of the price curve, since any power that we buy between two maxima can be sold at the second maximum. Similarly, the sell threshold switches at the minimum of the price curve, because any power that we sell between two minima can be bought at the first minimum.

The introduction of regimes impacts the parameter vector θ. Since we introduce different thresholds for the different regimes, μ_X and μ_Y are now indexed by the regime k, and are thus denoted as μ_X^k and μ_Y^k. In the remainder of this section, we will express these threshold means[5] μ_X^k and μ_Y^k as a function of 10 parameters, which we denote as $(\alpha_1^s,\alpha_1^b,\alpha_2^s,\alpha_2^b,\alpha_3^s,\alpha_3^b,\alpha_4^s,\alpha_4^b,\alpha_5^s,\alpha_5^b)$. We will then show how the REINFORCE algorithm can be used in order to learn the values of the parameter vector α.

B. Intraday auction curve

Our motivation for using the intraday auction curve as a feature for determining thresholds is illustrated in Fig. 5, where we present the *CIM price* for two different trading days. From this graph it is clear that it is not possible to set a single threshold which would perform well for both days, because the average level of the curves is different. In order to set an appropriate base level for the thresholds, we use the intraday auction price. The idea is that the price of previous markets can provide an indication about the state of the market, and thus support the forecast of the price for subsequent market-clearing stages. This observation has been inspired by: (i) reference [35], where the authors find a strong correlation between the day-ahead market and the balancing market; and (ii) reference [9], where the authors use the day-ahead price in order to forecast the imbalance price.

[5]In contrast to the mean, we do not make the standard deviation dependent on exogenous parameters. This is due to the fact that the standard deviation is only used in order to ensure sufficient exploration in the learning phase.

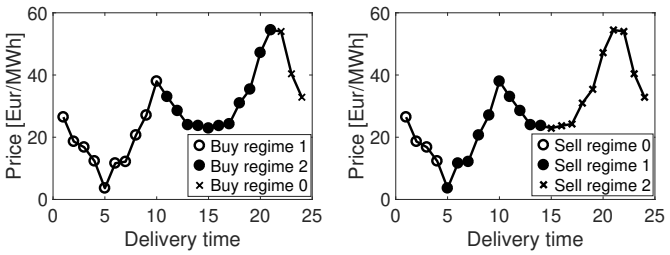


Fig. 4: Buy regime (left), and sell regime (right) based on the intraday auction price.

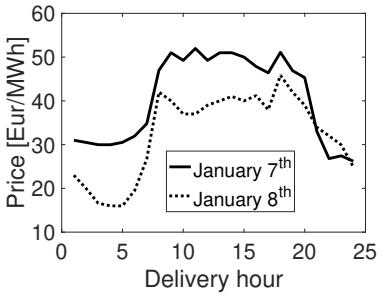


Fig. 5: Continuous intraday market price for two different days. The curves correspond to different average values, therefore different thresholds need to be applied for an effective threshold strategy.

Motivated by this observation, we propose an adaptation of the thresholds as follows:

$$\mu_X^k \leftarrow p_{\min,k} + \alpha_1^s(p_{\max,k} - p_{\min,k})$$
$$\mu_Y^k \leftarrow p_{\max,k} - \alpha_1^b(p_{\max,k} - p_{\min,k})$$

where $p_{\min,k}$ is the minimum of the k^{th} buy regime of the intraday auction curve, $p_{\max,k}$ is the maximum of the k^{th} sell regime of the intraday auction curve, and $\alpha_1^{s/b}$ are the weights that will be optimized using the REINFORCE algorithm.

The idea behind this parametrization is that $p_{\min,k}$ (resp. $p_{\max,k}$) is a reasonable starting point for the buy (resp. sell) threshold because it is the best price that could have been obtained in the intraday auction for regime k. Then, with the parameter α_1^s (resp. α_1^b) we allow the REINFORCE algorithm to determine to what extent the threshold should move from $p_{\min,k}$ (resp. $p_{\max,k}$) to $p_{\max,k}$ (resp. $p_{\min,k}$), based on learning from repeated episodes. In order to apply the REINFORCE algorithm for learning the parameter vector α, we need to

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

compute the derivative of our policy with respect to α. The derivative can be computed using the chain rule, as we show in Eq. (3) for the derivative of α_1^b, for delivery hour d in regime k.

$$\frac{\partial \pi_\theta^d(a|s)}{\partial \alpha_1^b} = \frac{\partial \pi_\theta^d(a|s)}{\partial \theta}^T \frac{\partial \theta}{\partial \alpha_1^b} = \frac{\partial \pi_\theta^d(a|s)}{\partial \mu_Y^k}(p_{\min,k} - p_{\max,k}) \quad (3)$$

C. Quantity already traded

The intuition for this adaptation is that, at any stage of the trading process, if we have already bought a large quantity of power and have not sold it yet, we wish to avoid the risk of ending up with unsold power. Note that we assume that there is no residual value for leftover water in the reservoir at the end of the horizon, which is consistent with the fact that we have an interest in entering a new day with an empty reservoir and filling the reservoir up with cheap power that is available during the night hours.

In order to capture this effect, we add a penalty in order to accept buying at a lower price and to accept selling at a lower price:

$$\mu_X^k \leftarrow \mu_X^k - \alpha_2^s \cdot v_{\text{end}}$$
$$\mu_Y^k \leftarrow \mu_Y^k - \alpha_2^b \cdot v_{\text{end}}$$

where v_{end} is the volume that we would obtain at the last delivery period with the trades that we have already engaged in. This adjustment of the thresholds implies that, moving forward, we become less selective about selling power and more selective about buying power, until the reservoir eventually becomes empty.

D. Remaining time before market closure

Whenever the producer has not sold all the energy stored in its reservoir close to the maximum of a regime, the producer should become less selective in the price it asks. This is due to the fact that there are few subsequent opportunities to trade, and the currently observed price is possibly the best price that the producer can secure for the trade. Similarly, whenever the producer has not bought up to the capacity of its reservoir as it is approaching the minimum of a regime, it should become less selective with the price that it asks for buying power. This approach is inspired by the theory of the optimal stopping problem [36], [37].

We capture this effect by varying the threshold means as follows:

$$\mu_X^k \leftarrow \mu_X^k + \alpha_3^s \frac{p_{\max,k} - p_{\min,k}}{2} \exp(\alpha_4^s(t - T_k^s)) \quad (4)$$

$$\mu_Y^k \leftarrow \mu_Y^k - \alpha_3^b \frac{p_{\max,k} - p_{\min,k}}{2} \exp(\alpha_4^b(t - T_k^b)) \quad (5)$$

where t is the current time step, T_k^b is the delivery time of the maximum of the k^{th} sell regime, and T_k^s is the delivery time of the minimum of the k^{th} buy regime.

We employ 4 coefficients in Eqs. (4) and (5): (i) $\alpha_3^{s/b}$ determines the strength of this effect; and (ii) $\alpha_4^{s/b}$ determines how smoothly the threshold adapts with respect to the gate closure time. A large value for $\alpha_4^{s/b}$ would decrease the selectivity very close to the delivery time. On the contrary, a small value for $\alpha_4^{s/b}$ would decrease the selectivity more smoothly with respect to the remaining time.

E. Relative value of observable bids

The motivation for this factor is to account for the coupling among the bids of different delivery hours, due to the fact that the battery can only store a finite amount of energy. Concretely, we wish to avoid accepting a bid even though the order book includes a bid at an adjacent delivery period that can be traded for a better price. In order to account for this inter-dependency, we penalize the bids that would not be accepted by the rolling intrinsic method. The rolling intrinsic policy is a myopic method for trading in continuous markets. This method has already been used as a benchmark in the literature [22], [23], and was originally proposed by [38] in the context of trading gas. The idea of the method is to trade so as to maximize the instantaneous reward at each time step [28]. In the context of our problem, the rolling intrinsic method will select the subset of trades which can be absorbed by the reservoir without exposing the unit to imbalances, and will do so by maximizing the profit of the current time step. This myopic policy can be written as an optimization problem at every time step of trade. The optimization model is developed in the electronic supplement.

Concretely, the adjustment to our algorithm is illustrated in Fig. 6. The figure corresponds to the case in which rolling intrinsic sells 20 MWh for delivery period d. When this occurs, we wish to decrease the probability of selecting the action of selling 30 MWh, and reallocate it to the probability of selling 20 MWh. To this end, we introduce an auxiliary Gaussian distribution with a mean of $\mu_Y^k + \alpha_5^b$ and with a standard deviation of $\exp(\sigma_Y)$. We compute the probability of the action *Sell* 30 MWh by using a threshold drawn from the auxiliary Gaussian distribution, which is indicated with the black bell curve in the figure. This decreases the probability of the action *Sell* 30 MWh, relative to the probability that would have been obtained from the original bell curve of Fig. 6. The difference in probability mass is transferred to the last action that is accepted by rolling intrinsic (*Sell* 20 MWh), as illustrated in Fig. 6. As we can see in the figure, the higher the value of α_5^b, the less likely we are to choose the action that is not selected by rolling intrinsic. The computation of the closed-form expression for $\pi_\theta^d(a|s)$ and its derivative is presented in detail in the electronic supplement.

F. Preventing imbalances

As we explain in the introduction, we are only interested in developing trading strategies that do not result in imbalance, given the amount of water that is currently stored in the reservoir. In order to remove actions that result in imbalances, we re-assign their probability to the closest action which does not result in an imbalance, using the same idea as in section

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

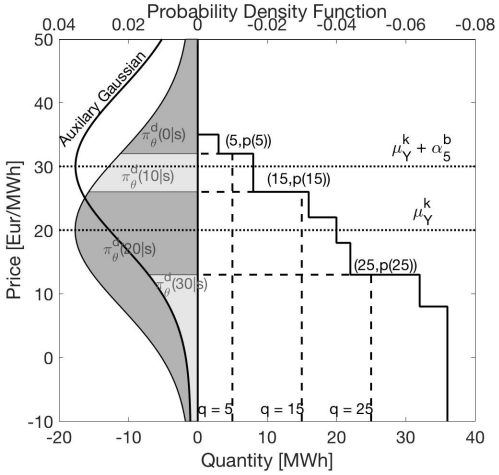


Fig. 6: Illustration of the probability reallocation that relies on the auxiliary Gaussian distribution, as described in section V-E.

V-E. In this case, the parameter α_5^b is replaced by a constant M which is sufficiently large in order to ensure that an action which would result in an imbalance is never selected.

G. Adapting with respect to round-trip efficiency

In order to account for round-trip efficiency, we present an example that illustrates the concept of *perceived* efficiency, which distinguishes whether we are planning to cover a bid *financially* or *physically*. Suppose that we have two delivery hours, a charging efficiency η_{in} of 0.9 and a discharging efficiency η_{out} of 0.9. Suppose that we have already bought 20 MWh for the first delivery hour at the previous time step. Therefore, the quantity that would be stored is 18 MWh for both delivery times. If we want to sell power at the second delivery time, we can only sell 16.2 MWh, because we have to apply the discharge efficiency. We define the *perceived* efficiency for this order as $\frac{16.2}{18} = 0.9$. On the contrary, if we want to sell at the first delivery time, we can sell 20 MWh, because this operation will simply cancel the previous purchase of 20 MWh. This is a purely financial operation. We thus define the *perceived* efficiency for this order as $\frac{20}{18} = 1.11$.

In order to account for this effect in the threshold parametrization, we use the same idea as in section V-E.

- We determine a certain *baseline* for the mean of the Gaussian distribution of our buy and sell threshold, which corresponds to the case in which we are accepting a certain quantity that serves as a purely *financial* transaction. We then adapt the thresholds as follows:

$$\mu_X^k \leftarrow \frac{1}{\eta_{\text{out}}}\mu_X^k$$
$$\mu_Y^k \leftarrow \eta_{\text{in}}\mu_Y^k$$

  This adaptation is coherent with the intuition presented in the example. Indeed, if we are canceling a position that we have previously taken in the market, we can accept a less interesting price (i.e. accepting a lower sell / higher buy threshold), because the perceived efficiency is higher than 1.
- The mean of the auxiliary Gaussian represents the case in which we are opening a new position. Therefore, the auxiliary Gaussian distribution mean will be equal to: (i) $\eta_{\text{in}} \cdot \eta_{\text{out}} \cdot \mu_X$ for the buy threshold; and (ii) $\frac{\mu_Y}{\eta_{\text{in}} \cdot \eta_{\text{out}}}$ for the sell threshold. This is also coherent with the example, because we are requesting a more selective price if we are opening a new position than if we are engaging in a purely financial transaction, since the *perceived efficiency* is less good. Note that this adaptation does not add any new parameters in the learning algorithm.

VI. Case study

In this section we present results from the implementation of the proposed policy on the German continuous intraday market. The data for the German CIM has been procured from the European Power Exchange (EPEX), and spans two years. For the purpose of the case study, we place ourselves in the position of a storage asset owner who manages a unit with a maximum storage capacity of 200 MWh. We assume that, on July 19, the owner adopts our strategy and has at its disposal market data since the beginning of the year[6]. Therefore, we use the 200 first days of 2015 as training set, and the last 165 days of 2015 and the 366 days of 2016 as a test set.

A. Learning process

We aim at learning the optimal threshold, so as to apply our threshold policy with a frequency of 1 second. We consider 1 second as a sufficiently high frequency for testing the algorithm in the continuous intraday market because, as observed in Table I, if we trade every second, we will observe 98.3% of the offers. This means that almost all of the offers remain in the market for at least one second, before being matched with competing offers on the platform.

In order for the learning stage of the algorithm to be computationally tractable, we gradually refine the learning frequency from hourly steps to 15-minute steps and ultimately to 5-minute steps[7], as indicated in Fig. 7. In this figure, 1 iteration corresponds to 4 repetitions of the 200 days of learning, which amounts to 800 episodes. These episodes are executed in parallel on an HPC cluster using 8 CPUs for 40 hours[8].

A potential issue for our learning phase is that the REINFORCE algorithm is a stochastic algorithm. Therefore, different runs can produce different results. In order to test the sensitivity of our results, we have conducted an experiment

[6]Note that data which extends too far back in time may not be as useful, due to the rapid structural evolution of the market (increase in renewable energy integration, changes in market design, etc.).

[7]We decide to switch to a higher learning frequency when the profit appears to stabilize. This is due to the fact that there is no reason to run the algorithm until full convergence for the hourly frequency, because it is not the problem we are interested in (the order data arrives at much higher frequency than hourly).

[8]We have performed an analysis on the computational scalability of our learning process. The main message of the study is that the run time increases linearly with respect to the trading frequency. The complete numerical analysis has been added in the electronic supplement.

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

Length of time step	Percentage of offers observed
1 hour	25.7
15 minutes	41.5
5 minutes	56
1 minute	74.8
15 seconds	86.7
5 seconds	92.2
1 second	98.3

TABLE I: Percentage of offers that are observed as a function of frequency of accessing the market data.

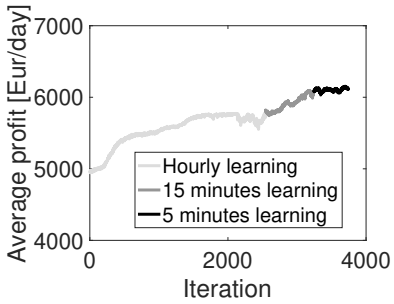


Fig. 7: Evolution of profit as a function of iterations of the REINFORCE algorithm.

in which we run 6 different realizations of the REINFORCE algorithm at hourly frequency and compare the evolution of the alpha parameters and the profit. We find that the different runs exhibit a very similar performance. The complete experiment is available in the electronic supplement of the paper.

B. Out-of-sample testing

In this section, we present the results obtained by our threshold policy on out-of-sample data. More precisely, we apply the θ parameter vector learned on the 200 first days of 2015 on the remainder of 2015 and to the entire year of 2016. We compare these results with the ones obtained by the rolling intrinsic method, which is described in section V-E. . We present the results in Table II. (i) Column 1 represents the trading frequency. (ii) Column 2 refers to the method that we use: "Threshold" refers to the method developed in this paper, "Rolling" refers to rolling intrinsic, "GM" is the policy that has been developed in [23] and "Threshold without α_i" is the policy learned by the REINFORCE algorithm if we fix α_i^s and α_i^b to 0. (iii) Column 3 refers to the round-trip efficiency of the considered storage unit. (iv) Column 4 refers to the data that are used for the test. It can either be in-sample (the 200 first days of 2015) or out-of-sample (the remainder of 2015 and 2016). (v) Column 5 contains the average profit. From this table, 5 main observations can be made: (i) our threshold policy outperforms rolling intrinsic; (ii) our threshold policy is more suited for high frequency than the one presented in [23]; (iii) our threshold policy also outperforms rolling intrinsic for a non perfect round-trip efficiency; (iv) the results in and out-of-sample are very similar; and (v) the most important parameters are α_3 α_4 and α_5. In the remainder of this section, we will analyse these five observations in more details.

*a) **Superiority of the threshold policy compared to rolling intrinsic**:* By observing rows 4 and 5 of the table,

Trading frequency	Method	Efficiency	Used data	Profit mean
1 hour	Threshold	1	out	5374
1 hour	Rolling	1	out	4591
1 hour	GM	1	out	4776
1 second	Threshold	1	out	6405
1 second	Rolling	1	out	5438
1 second	GM	1	out	5186
1 second	Threshold	0.81	out	3762
1 second	Rolling	0.81	out	3311
1 second	Threshold	1	in	6605
1 second	Rolling	1	in	5694
1 hour	Threshold without α_1	1	out	5362
1 hour	Threshold without α_2	1	out	5375
1 hour	Threshold without α_3 and α_4	1	out	4652

TABLE II: Profit mean [€/day]

we observe that the average profit difference amounts to 17.8%. Moreover, the proposed threshold policy achieves a higher profit in 77.4% of the days. In Fig. 8, we present the daily profit difference between the threshold and the rolling intrinsic policy. The figure demonstrates that the extra profit is a cumulative effect of multiple days of superior performance, as opposed to being the result of a few isolated days in which the threshold policy performed significantly better.

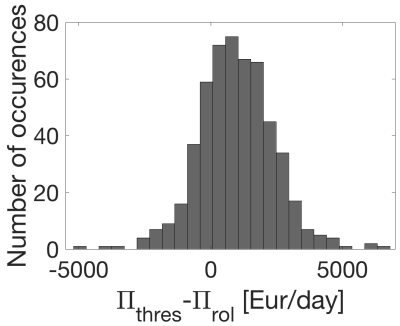


Fig. 8: Distribution of the difference between the profit of the threshold policy and the rolling intrinsic policy

In Fig. 9, we illustrate one of the effects that justifies the superior profit of the threshold policy. This graph indicates whether power has been traded for the different delivery times and time steps. An empty dot indicates that we have bought power, whereas a solid dot indicates that we have sold power. The left graph illustrates one of the weaknesses of rolling intrinsic: at each time step where there are empty dots, there are also solid dots. This implies that the method only considers buying power if it can sell it directly (except if it can buy power at a negative price). This is due to the fact that the method maximizes the profit of the current time step, and ignores future trading opportunities which may arrive but have not yet been observed. On the contrary, the threshold method procures power at the beginning of the horizon, but may turn down offers for selling power if the sales price is not sufficiently attractive. Thus, the threshold method may wait in order to sell the power later, counting on the possibility that at a later moment there will be offers arriving in the market

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

with a higher willingness to pay than the currently available offers.

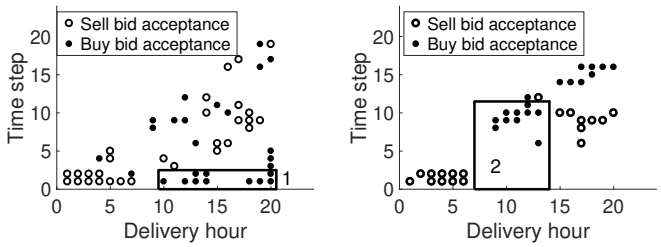


Fig. 9: Bid acceptance patterns for 1 day of trading for the rolling intrinsic (left) and threshold method (right).

b) ***Comparison of our threshold policy and the GM policy at high frequency****:* We compare the influence of the trading frequency on the performance of three different methods: (i) the threshold policy presented in the present paper; (ii) the rolling intrinsic policy; and (iii) the GM policy. Note that the GM policy does not incorporate the parameters α_5^s and α_5^b in the parametrization of the threshold. From rows 1-6 of the table, we observe that the profit increases with respect to the trading frequency for all the methods. This is expected, since an increase in the trading frequency increases the number of offers that we observe and use for trading. However, this profit increase is smaller for the GM policy. In order to interpret this result, Figs. 10 and 11 compare the evolution of five different policies: (i) rolling intrinsic with an hourly trading frequency; (ii) rolling intrinsic with a trading frequency of 15 seconds; (iii) rolling intrinsic with a trading frequency of 1 second; (iv) a threshold policy with a trading frequency of 1 second; and (v) the GM policy with a trading frequency of 1 second.

In Fig. 10, we observe that two factors contribute to the profit. *(i)* The first factor is the profit that results from the significant arbitrage possibilities of the storage unit. These arbitrage opportunities can be anticipated. These profits correspond to the large jump of the rolling intrinsic method. *(ii)* The second factor corresponds to the profits that result from trades of smaller volume, which are not visible at the outset of the trading day. These profits correspond to the small increase of the profit of the rolling intrinsic policy, following the large jump. It is worth noting that these small increases are almost insignificant when trading at an hourly time step, but become very important at a higher trading frequency.

This analysis highlights that, when trading at a higher frequency, we require a trading strategy that is effective at capturing the value of both predictable large arbitrage opportunities and less predictable small opportunities. In Fig. 11 we observe that the threshold policy attains similar performance to rolling intrinsic in terms of capturing small arbitrage opportunities. This is represented by the right graph, where we observe that the two curves follow a similar pattern towards the end of the day. On the contrary, the GM policy is not able to capture these small arbitrage opportunities, which is clear from the fact that the profit remains constant at the end of the day. Note that the GM policy parametrization does not include any information about the prices available for the other delivery hours. The problem is that, for these small arbitrage opportunities, a bid is not interesting only due to its price but also because if we accept it along with a bid with another delivery time, we can directly secure a positive profit using our storage unit.

On the other hand, the main difference between our threshold policy and rolling intrinisc mainly rests on the fact that the threshold policy is better suited for trading for big arbitrage opportunities. This is illustrated by the fact that the large jump of the threshold policy is higher than that of rolling intrinsic. The rolling intrinsic policy buys and sells prematurely in the beginning of the day, whereas the threshold policy holds back until more favorable trades can be locked in.

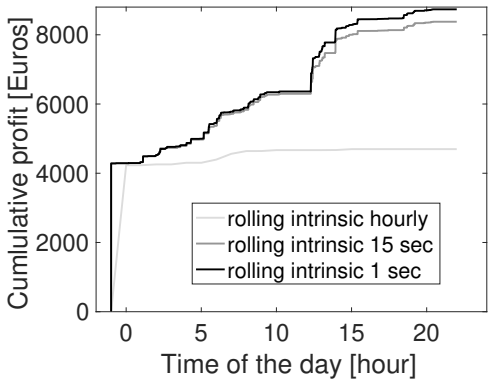


Fig. 10: Profit evolution for one day of trading for the rolling intrinsic method for various trading frequencies.

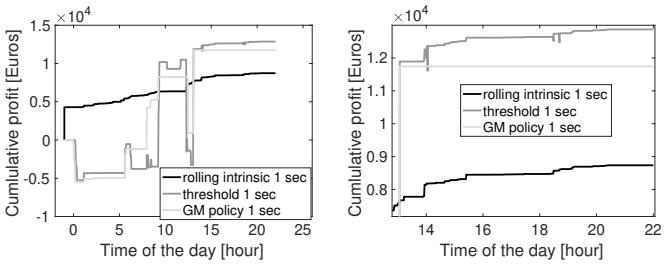


Fig. 11: Profit evolution for one day of trading for the full day (left) and a zoom in on the end of the day (right).

c) ***Threshold performance for a non perfect roundtrip efficiency****:* In this section, we present the results for a storage unit with a charging efficiency of 0.9 and a discharging efficiency of 0.9. Our aim is to verify that our threshold policy is also suitable for an asset with an imperfect round-trip efficiency. The results are presented in rows 7 and 8 of the table. As before, we compare the results obtained by our threshold policy with the ones obtained by rolling intrinsic on the same data. The proposed threshold policy achieves a higher profit in 64.6% of the days. The average profit difference amounts to 13.6%. In Fig. 12 we present the daily profit difference. These results are relatively close to the ones obtained for a storage unit with a perfect round-trip efficiency. The results thus suggest that our policy is also suitable for the case with round-trip efficiency losses.

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

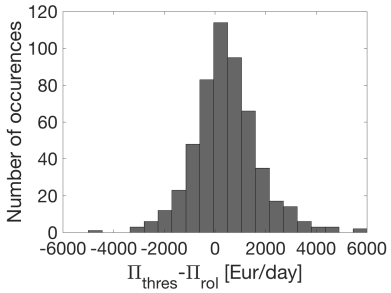


Fig. 12: Distribution of the difference between the profit of the threshold policy and the rolling intrinsic policy in the case with round-trip efficiency losses.

*d) **Stability of the method with respect to change in the data**:* In rows 5 and 10 of the table, we compare the profit obtained by rolling intrinsic in-sample and out-of-sample. We observe that the profit is slightly higher in-sample than out-of-sample. Thus, the performance under in-sample data is slightly more favourable than under out-of-sample data. In rows 4 and 9, we observe that our threshold policy also achieves a slightly higher profit in-sample. Note that the difference between the two methods amounts to 911 € in-sample and 966 € out-of-sample. Thus, our method is observed to achieve a robust performance against out-of-sample data.

*e) **Importance of the different parameters**:* In order to test the influence of each element of the threshold parametrization, we have launched the REINFORCE algorithm by cancelling each of the parameters one by one. Then we apply the learned policy out-of-sample. More precisely, we have realized 3 simulations: (i) we optimize the policy while fixing α_1^s and α_1^b to 0; (ii) we optimize the policy while fixing α_2^s and α_2^b to 0; and (iii) we optimize the policy while fixing α_3^s, α_3^b, α_4^s and α_4^b to 0. The reason for fixing both α_3 and α_4 together is that cancelling one will cancel the other automatically.

We have not considered the case in which we set α_5^s and α_5^b to 0, because the importance of this parameter is already discussed extensively when analysing the GM policy. The results are presented in the last three rows of Table II and will be analysed in detail in the remainder of the section.

1) **Without α_1^s and α_1^b:** We observe that the profit is slightly better if we include α_1^s and α_1^b compared to the case in which we set them to 0. This suggests that these parameters help, but their impact is not decisive[9].
2) **Without α_2^s and α_2^b:** We observe that the profit with and without α_2^s and α_2^b is almost identical. This suggests that this parameter could be discarded without hurting the profitability of the policy.
3) **Without α_3^s, α_3^b, α_4^s and α_4^b:** In this case, we observe a big drop in the profit, compared to the initial case. By further investigating the obtained parameters, we observe that the algorithm converges to a high value of α_1^s and α_1^b. This implies that the sell threshold will be low and the buy threshold will be high. Simultaneously, the algorithm increases the parameters α_5^s and α_5^b to a very high value. The consequence of this behaviour will be that the policy will aim at accepting every possible bid that is also accepted by rolling intrinsic. This indicates that this policy is attempting to mimic the rolling intrinsic policy. In order to confirm this intuition, we present in Fig. 13 the profit difference between the rolling intrinsic policy and this policy. We observe that the values are concentrated around 0, which confirms our intuition that the algorithm is attempting to mimic rolling intrinsic.

[9]Notice that we have realized the same experiment at a frequency of 1 minute, and the extra profit was more significant (around 0.83%).

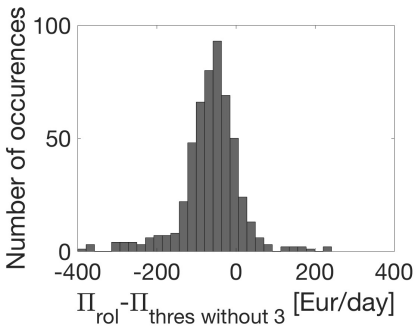


Fig. 13: Distribution of the difference between the profit of the rolling intrinsic policy and the profit of the threshold policy without parameters $\alpha_3^{s/b}$ and $\alpha_4^{s/b}$

VII. Conclusions and perspectives

In this paper we tackle the problem of intraday trading for storage units. We model the problem using Markov Decision Processes. We focus on policies that are parametrized on price thresholds, and we optimize the resulting policy functions using the REINFORCE algorithm. We introduce and justify a collection of factors that can be used for adapting the trading threshold to system conditions. We compare our threshold policy to the rolling intrinsic method on the German continuous intraday market. We demonstrate that the threshold policy performs significantly better than rolling intrinsic, and analyze the results in order to explain the performance difference. In future work, we are interested in improving the policy functions by adding more explanatory variables of the price thresholds such as renewable forecasts or generator outages. We also aim at developing trading strategies in the Continuous Intraday Market for renewable sources with uncertain real-time production.

Acknowledgment

This research was funded by the ENGIE Chair on Energy Economics and Energy Risk Management, by an Electrabel grant on "Modeling the Value of Flexibility at Sub-Hourly Operating Time Scales", and by the Belgian National Science Foundation (FSR-FNRS) through a FRIA grant. The authors also want to thank Gauthier de Maere and Guillaume Erbs for their helpfull comments during the development of this work.

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2957246, IEEE Transactions on Power Systems

REFERENCES

[1] EU [Online]. Available: https://ec.europa.eu/clima/policies/strategies/2020_en

[2] Eurostat [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/images/d/de/Table_3-Share_of_electricity_from_renewable_sources_in_gross_electricity_consumption_2004-2016.png

[3] EPEXSPOT [Online]. Available: https://www.epexspot.com/fr/donnees_de_marche/intradaycontinuous

[4] J. Matevosyan and L. Soder, "*Minimization of imbalance costs trading wind power on the short-term power market,*" in IEEE Transaction on Power Systems, vol. 21, no. 3, pp. 1396-1404, Aug. 2006.

[5] J. Chaves, R. Hakvoort and A. Ramos, "*Short-term strategies for Dutch wind power producers to reduce imbalance costs,*" in Energy Policy, vol. 52, pp. 573-582, Jan. 2013.

[6] D. Lee, H. Shin, and R. Baldick, "*Bivariate Probabilistic Wind Power and Real-Time Price Forecasting and Their Applications to Wind Power Bidding Strategy Development,*" in IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6087-6097, Nov. 2018.

[7] C. O'Dwyer and D. Flynn, "*Using Energy Storage to Manage High Net Load Variability at Sub-Hourly Time-Scales,*" in IEEE Transactions on Power Systems, vol. 30, no. 4, pp. 2139-2148, July 2015.

[8] M. Khodayar, M. Shahidehpour, and L. Wu, "*Enhancing the Dispatchability of Variable Wind Generation by Coordination With Pumped-Storage Hydro Units in Stochastic Power Systems,*" in IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 2808-2818, Aug. 2013.

[9] T. Boomsma, N. Juul and S. Fleten, "*Bidding in sequential electricity markets: The Nordic case,*" in European Journal of Operational Research, vol. 238, no. 3, pp.797-809, Nov. 2014.

[10] S. Braun, "*Hydropower Storage Optimization Considering Spot and Intraday Auction Market,*" in Energy Procedia, vol. 87, pp. 36-44, Jan. 2016.

[11] R. Kiesel and F. Paraschiv, " *Econometric analysis of 15-minute intraday electricity prices,*" in Energy Economics, vol. 64, pp. 77-90, May 2017.

[12] F. Ziel, "*Modeling the impact of wind and solar power forecasting errors on intraday electricity prices,*" 2017 14th International Conference on the European Energy Market (EEM), Dresden, 2017.

[13] C. Balardy, "*An Empirical Analysis of the Bid-Ask Spread in the German Power Continuous Market,*" working paper, 2018.

[14] R. Kiesel. "*Modeling market order arrivals on the intraday power market for deliveries in Germany with Hawkes processes with parametric kernels,*" Energy Finance Christmas Workshop, 2017.

[15] H. Martin and S. Otterson, "*German Intraday Electricity Market Analysis and Modeling Based on the Limit Order Book,*" 2018 15th International Conference on the European Energy Market (EEM), Lodz, 2018.

[16] S. Braun and R. Hoffmann, "*Intraday Optimization of Pumped Hydro Power Plants in the German Electricity Market,*" in Energy Procedia vol. 87, pp. 45-52, Jan. 2016.

[17] E. Engmark, H. Sandven, S. Fleten and G. Klaboe, "*Stochastic Multi-stage Bidding Optimisation in an Intraday Market with Limited Liquidity,*" 2018 15th International Conference on the European Energy Market (EEM), Lodz, 2018.

[18] E. Garnier and R. Madlener, "*Balancing forecast errors in continuous-trade intraday markets,*" in Energy Systems, vol. 6, no. 3, pp. 361–388, Sep. 2015.

[19] R. Aid, P. Gruet and H. Pham, "*An optimal trading problem in intraday electricity markets,*" in Mathematics and Financial Economics, vol. 10, no. 1, pp. 49-85, Jan. 2016.

[20] A. Skajaa, K. Edlund and J. Morales, "*Intraday Trading of Wind Energy,*" in IEEE Transactions on Power Systems, vol. 30, no. 6, pp. 3181-3189, Nov. 2015.

[21] G. Bertrand and A. Papavasiliou, "*An Analysis of Threshold Policies for Trading in Continuous Intraday Electricity Markets,*" 2018 15th International Conference on the European Energy Market (EEM), Lodz, 2018.

[22] I. Boukas et al., "*Intra-day Bidding Strategies for Storage Devices Using Deep Reinforcement Learning,*" 2018 15th International Conference on the European Energy Market (EEM), Lodz, 2018.

[23] G. Bertrand and A. Papavasiliou, "*Reinforcement-Learning Based Threshold Policies for Continuous Intraday Electricity Market Trading,*" IEEE PES General Meeting, Atlanta, 2019.

[24] EPEXSPOT [Online]. Available: https://www.epexspot.com/en/product-info/auction/germany-luxembourg

[25] EPEXSPOT [Online]. Available: https://www.epexspot.com/en/product-info/intradayauction/germany

[26] EPEXSPOT [Online]. Available: https://www.epexspot.com/en/product-info/intradaycontinuous/germany

[27] TENNET [Online]. Available: https://www.tennet.eu/fileadmin/user_upload/The_Electricity_Market/German_Market/Grid_customers/contracts/BNetzA-BKC_englisch.pdf

[28] W. Powell and S. Meisel, "*Tutorial on Stochastic Optimization in Energy—Part II: An Energy Storage Illustration,*" in IEEE Transactions on Power Systems, vol. 31, no. 2, pp. 1468-1475, Mar. 2016.

[29] R. Sutton and A. Barto, "*Reinforcement Learning: An Introduction,*" in MIT press, 2018.

[30] R. J. Williams, "*Simple statistical gradient-following algorithms for connectionist reinforcement learning,*" in Machine Learning, vol. 8, no. 3-4, pp. 229-256, May 1992.

[31] W. Morris, "*Some analysis of purchasing policy,*" in Management Science, vol. 5, no.4, pp. 443-452, Jul. 1959.

[32] B. Kingsman, "*Commodity purchasing,*" in Journal of the Operational Research Society, vol. 20, no. 1, pp. 59-79, Mar. 1969.

[33] K. Golabi. "*Optimal inventory policies when ordering prices are random,*" in Operations Research, vol. 33, no. 3, 1985.

[34] W. Powell and S. Meisel, "*Tutorial on Stochastic Optimization in Energy-Part I:Modeling and Policies,*" in IEEE Transactions on Power Systems vol. 31, no. 2, pp. 1459-1467, Mar. 2016.

[35] G. Klaboe, A. Eriksrund and S. Fleten, "*Benchmarking time series based forecasting models for electricity balancing market prices,*", in Energy Systems, vol. 6, no. 1, pp. 43–61, Mar. 2015.

[36] D. Lindley, "*Dynamic programming and decision theory,*" in Applied Statistic, vol. 10, no.1, pp. 39-52, 1961.

[37] J. Tsitsiklis and B. Van Roy "*Optimal Stopping of Markov Processes: Hilbert Space Theory, Approximation Algorithms, and an Application to Pricing High-Dimensional Financial Derivatives,*" in IEEE Transactions On Automatic Control, vol. 44, no. 10, Oct. 1999.

[38] J. Gray and P. Khandelwal, "*Realistic natural gas storage models II: Trading strategies,*" in Commodities Now, Sep. 2004.



Gilles Bertrand (M' 16) obtained his B.Sc. and M.Sc. degrees in Mathematical Engineering from UCLouvain, Louvain-la-Neuve, Belgium. He is currently a Ph.D. student on Applied Mathematics at the Center for Operations Research and Econometrics of UCLouvain, Louvain-la-Neuve, Belgium.



Anthony Papavasiliou (M' 06) received the B.S. degree in electrical and computer engineering from the National Technical University of Athens, Greece, and the Ph.D. degree from the Department of Industrial Engineering and Operations Research (IEOR) at the University of California at Berkeley, Berkeley, CA, USA. He holds the ENGIE Chair at UCLouvain, Belgium.

0885-8950 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.