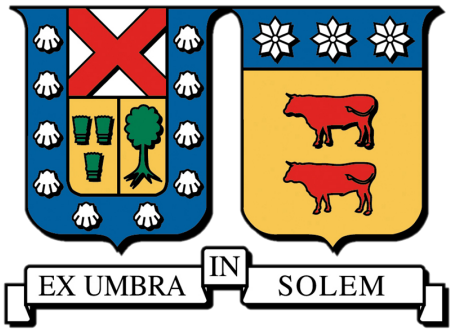


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

Multi-agent Deep Reinforcement Learning for Efficient Multi-Timescale Bidding of a Hybrid Power Plant in Day-Ahead and Real-Time Markets

---

Tomás Ochoa Abett de la Torre

Thesis submitted in partial fulfillment of the requirements for the degree of
MAGISTER EN CIENCIAS DE LA INGENIERÍA ELÉCTRICA

Thesis Director: Dr. Esteban Gil

Valparaíso, Julio 2022

Multi-agent Deep Reinforcement Learning for Efficient Multi-Timescale Bidding of a Hybrid Power Plant in Day-Ahead and Real-Time Markets

Tomás Ochoa Abett de la Torre

Tesis presentada al Comité de Evaluación
de Tesis integrado por:

Director de Tesis: **Esteban Gil**
Presidente del Comité: **Alejandro Angulo**
Co-referente Externo: **Carlos Valle**

Para completar las exigencias del grado de
Magíster en Ingeniería Eléctrica

Departamento de Ingeniería Eléctrica
Universidad Técnica Federico Santa María
Valparaíso, Chile
Julio 2022

"My reality needs imagination like a bulb needs a socket. My imagination needs reality like a blind man needs a cane." **- Tom Waits**

Acknowledgments

To begin with, I would like to thank my mother and father for always giving me the necessary means to carry out my studies and, together with my brother, for giving me loving, support, and advice whenever I needed it.

I thank those friends who have been around to share all kinds of conversations and experiences. They have helped me to become the person I am now and gave me all the needed to construct my life path. I have no doubts I will be trying my best to keep these human links ever long.

I also would like to thank and highlight, the tremendous academic experience and role modeling examples that I have received from professors Esteban Gil, Alejandro Angulo, and Carlos Valle. This research work has been majorly boosted by their participation throughout its elaboration. Special thanks to professor Esteban Gil, who has been a guide for me and has opened the door to pursue this degree.

The Universidad Técnica Federico Santa María has been my second home for the last few years. It has given me a place where I have managed to challenge myself and clarify what must be my priorities. Thanks for the support given by the Chilean National Agency for Research and Development (ANID) through the grants 1210625 and FB0008, and by the DPP of Universidad Técnica Federico Santa María under grants PI-LIR-2020-59 and PIIC-005/2021.

Contents

List of Figures

List of Tables

Resumen

La oferta eficiente de múltiples productos eléctricos en condiciones de incertidumbre permitiría una participación de mercado más rentable para las centrales eléctricas híbridas con recursos energéticos variables y sistemas de almacenamiento, así ayudando al proceso de descarbonización. Este estudio trata sobre la licitación eficiente de una planta fotovoltaica con un sistema de almacenamiento de energía (PV-ESS) que participa en mercados eléctricos de múltiples escalas temporales, proporcionando productos de energía y servicios auxiliares (AS). El sistema de gestión de energía (EMS) tiene como objetivo maximizar las ganancias de la planta mediante una oferta eficiente en los mercados diarios y de tiempo real, considerando la entrega adecuada de los productos adjudicados. Las decisiones de licitación del EMS generalmente se obtiene usando métodos de optimización tradicionales. Sin embargo, dado que el problema abordado es un programa estocástico de múltiples etapas, a menudo el problema es intratable y sufre *curse of dimensionality*. Este documento presenta un método novedoso consistente en aprendizaje profundo reforzado multiagente (MADRL) para la licitación eficiente a múltiples escalas de tiempo. Dos agentes basados en redes neuronales artificiales de vista múltiple con capas recurrentes (MVANN) se ajustan para mapear las observaciones del entorno en acciones. Dichos mapeos utilizan como entradas la información disponible relacionada con los productos del mercado eléctrico, las decisiones de licitación, la generación solar, la energía almacenada y las representaciones de tiempo para ofertar en ambos mercados eléctricos. Sostenido por una suposición de *price taker*, el entorno del EMS, el cual se encuentra limitado física y financieramente, se simula empleando datos históricos. Se utiliza una función de recompensa acumulativa compartida con un horizonte de tiempo finito para ajustar los pesos de ambas MVANNs simultáneamente durante la fase de aprendizaje. Se ha comparado el método MADRL propuesto con métodos de optimización estocásticos y robustos de dos etapas basados en escenarios. Los resultados se proporcionan para la participación de la planta híbrida durante un año usando una resolución de 1 minuto. El método propuesto logró mayores ganancias estadísticamente significativas, menos variabailidad en ingresos en ambos mercados eléctricos y una mejor provisión de los productos adjudicados al lograr desequilibrios energéticos más pequeños y menos variables a lo largo del tiempo.

Abstract

Effective bidding on multiple electricity products under uncertainty would allow a more profitable market participation for hybrid power plants with variable energy resources and storage systems, therefore aiding the decarbonization process. This study deals with the effective bidding of a photovoltaic plant with an energy storage system (PV-ESS) participating in multi-timescale electricity markets by providing energy and ancillary services (AS) products. The energy management system (EMS) aims to maximize the plant's profits by efficiently bidding in the day-ahead and real-time markets while considering the awarded products' adequate delivery. EMS's bidding decisions are usually obtained from traditional mathematical optimization frameworks. However, since the addressed problem is a multi-stage stochastic program, it is often intractable and suffers the curse of dimensionality. This document presents a novel multi-agent deep reinforcement learning (MADRL) framework for efficient multi-timescale bidding. Two agents based on multi-view artificial neural networks with recurrent layers (MVANNs) are adjusted to map environment observations to actions. Such mappings use as inputs available information related to electricity market products, bidding decisions, solar generation, stored energy, and time representations to bid in both electricity markets. Sustained by a price-taker assumption, the physically and financially constrained EMS's environment is simulated by employing historical data. A shared cumulative reward function with a finite time horizon is used to adjust both MVANNs' weights simultaneously during the learning phase. We compare the proposed MADRL framework against scenario-based two-stage robust and stochastic optimization methods. Results are provided for one-year-round market participation of the hybrid plant at a 1-minute resolution. The proposed method achieved statistically significant higher profits, less variable incomes from both electricity markets, and better provision of awarded products by achieving smaller and less variable energy imbalances through time.

Nomenclature

e_t^{s}	ESS's stored energy at minute t, `MWh`
$\hat{e}^{\mathrm{s}}$	ESS's maximum storage capacity, `MWh`
$\check{e}^{\mathrm{s}}$	ESS's minimum storage capacity, `MWh`
η^{d}	ESS's discharge efficiency, `-`
η^{c}	ESS's charge efficiency, `-`
p_t^{d}	ESS's off-terminal discharge power flow at minute t, `MW`
p_t^{c}	ESS's off-terminal charge power flow at minute t, `MW`
$\hat{p}^{\mathrm{s}}$	ESS's off-terminal rated power, `MW`
p_t^{pv}	PV's power flow at minute t, `MW`
p_t^{g}	PV-ESS's power flow to the grid at minute t, `MW`
p_t^{r}	PV-ESS's control reference signal at minute t, `MW`
δ_t^+	PV-ESS's under-generation at minute t, `MW`
δ_t^-	PV-ESS's over-generation at minute t, `MW`
p_h^{DA}	EMS's DA energy bid at hour interval h, `MW`
p_q^{RT}	EMS's RT energy bid at 15-min interval q, `MW`
p_h^{Ru}	EMS's DA capacity for up-regulation bid at hour interval h, `MW`
p_h^{Rd}	EMS's DA capacity for down-regulation bid at hour interval h, `MW`
$\hat{\alpha}$	EMS's upper bounds for energy products bids, `MW`
$\check{\alpha}$	EMS's lower bounds for energy products bids, `MW`
$\hat{\beta}$	EMS's upper bounds for capacity for up-regulation product bids, `MW`
$\hat{\gamma}$	EMS's upper bounds for capacity for down-regulation product bids, `MW`
λ_h^{DA}	DA energy product price at hour interval h, `$/MWh`
λ_q^{RT}	RT energy product price at 15-min interval q, `$/MWh`
λ_h^{Ru}	DA capacity for up-regulation product price at hour interval h, `$/MWh`
λ_h^{Rd}	DA capacity for down-regulation product price at hour interval h, `$/MWh`
b_t^+	ISO's signal for up-regulation deployment at minute t (scaled), `-`
b_t^-	ISO's signal for down-regulation deployment at minute t (scaled), `-`
λ^{imb}	Imbalance penalization value, `$/MWh`
Δ^{k}	Conversion factors for k minute-intervals, `h`
a_t^1	DA-MVANN action at minute t
a_t^2	RT-MVANN action at minute t
r_t	DA-MVANN and RT-MVANN shared reward signal at minute t
o_t^1	DA-MVANN observed state at minute t
o_t^2	RT-MVANN observed state at minute t
s_t	Environment state at minute t

1 Introduction

The variability and uncertainty of photovoltaic (PV) generation pose many challenges for integrating variable renewable energy sources onto existing electrical grids. Potential adverse effects on reliability and stability of electrical networks could limit their integration, as a higher penetration would increase frequency control requirements [1]. A potential solution to counteract a PV plant's naturally oscillating power output is to incorporate an energy storage system (ESS), resulting in a hybrid PV-ESS plant with the ability to shift energy injections and consumption through time and even provide frequency control capacity. An adequately controlled PV-ESS plant can provide electricity products traditionally provided by fossil fuel-based power plants, therefore aiding to decarbonize the electricity sector.

Different electricity products (e.g., energy and capacity for regulation) are valued through time in different markets, such as the day-ahead (DA) and real-time (RT) markets. Price signals indirectly report to market stakeholders the shortage or abundance in the supply of specific electricity products. Proper management of a PV-ESS plant would allow a more profitable participation in electricity markets by efficiently deciding on which time and market to allocate the plant's resources. Furthermore, different studies [2, 3] have shown that profits solely from energy arbitrage may be insufficient to recover the totality of the ESS's capital expenditures. However, these studies show that offering both energy and ancillary service products in different markets can boost their competitiveness.

The present document is an extended version of the article *Multi-agent Deep Reinforcement Learning for Efficient Multi-Timescale Bidding of a Hybrid Power Plant in Day-Ahead and Real-Time Markets* to be published on Applied Energy (Elsevier), currently in press. Appendix A lists articles generated in recent years as a master's student, related and not to the current research.

1.1 Hypothesis

A profit optimization scheme for a PV/ESS plant participating on DA, RT, and AS electricity markets based on MADRL for effective bidding can achieve competitive results against state-of-the-art control optimization methods such as stochastic and robust scenario-based MPC.

1.2 Research objectives

1.2.1 Main objective

The main objective of this research is to introduce a novel MADRL method for the efficient multi-timescale bidding of a hybrid PV/ESS power plant participating on

DA and RT electricity markets by providing energy and AS products. [1].

1.2.2 Specific objectives

1. Design, develop, and implement a MADRL application for efficient bidding of a hybrid PV/ESS plant participating on DA and RT electricity markets.
2. Design a two-stage scheme for the efficient bidding of a hybrid PV/ESS plant using scenario-based stochastic and robust optimization frameworks.
3. Test and compare the performance of stochastic and robust optimization scenario-based schemes against the proposed MADRL framework using out-of-sample historical data.

1.3 Structure of the document

The remainder of this document is organized as follows:

- Chapter II presents discusses the state-of-the-art regarding bidding in electricity markets and the use of machine learning techniques for sequential decision making.
- Chapter III provides background for the market structure, the PV-ESS system, and multi-time scale bidding in the DA and RT markets.
- Chapter IV presents the MADRL framework for solving the multi-stage problem in a staggered manner.
- Chapter V develops the proposed methodology for a case study and provides numerical results.
- Chapter VI reports the conclusions regarding the current research.

[1] In the Thesis Proposal document, the approach was stated as "rolling horizon two-stage ANN" instead of "MADRL" and "joint optimization and operation" was used instead of "efficient bidding" for problem description. These terms are nearly equivalent; nevertheless, Applied Energy reviewers recommended more clarity on how we were referring to these aspects, and therefore these have been updated in this document.

2 State of the art

In the last decade, bidding optimization for hybrid power plants with storage participating in multiple electricity markets has received much attention. In [4], a compressed air energy storage unit optimizes its bidding in the DA and RT markets, offering energy and reserves, but their deterministic optimization approach ignores the market price uncertainties faced by the plant. Deterministic approaches are often inadequate, as efficiently managing an ESS is a multi-stage stochastic optimization problem. However, multi-stage stochastic formulations are impractical since they are, in general, intractable and suffer the curse of dimensionality using conventional optimization frameworks [5]. Thus, the underlying problem is commonly approximated by two-stage stochastic or robust programming formulations [6, 7, 8, 9, 10, 11].

Previous works, such as [6], have used two-stage stochastic programming for the bidding of energy and spinning reserves in the DA market for a hybrid portfolio consisting of thermal and wind generation and compressed air energy storage. The problem maximizes the expected profits while simultaneously handling the risk by adding a Conditional Value at Risk (CVaR) term to the objective function. However, instead of explicitly modeling bidding in the RT market, a penalization is included for DA commitment deviations. Other authors also manage risk by adding a CVaR term into their formulations [7]. Work done in [8] proposes a two-stage scenario-based stochastic model to enable a hybrid power plant (wind-ESS) to participate in simultaneous day-ahead energy, spinning reserve, and frequency regulation markets under different operation strategies. Nevertheless, this work does not address the problem of multi-timescale bidding, as all markets operate once a day. In [9], the authors propose a two-stage robust optimization procedure (non-scenario-based) for a virtual power plant, which establishes confidence bounds for the uncertainty set. Reference [10] compares risk-neutral and risk-averse strategies employing two-stage scenario-based stochastic-robust programming for market participation in DA and RT markets of a hybrid charging station with a PV system. Usually, robust approaches yield conservative solutions since they are intrinsically designed to be sub-optimal, aiming to maximize the profits for extreme scenarios [12]. A drawback of traditional stochastic optimization approaches is their dependence on the quality of the uncertainty representation. Previous works, such as [11], have put much effort into refining scenario generation processes, aiming to improve the performance of stochastic bidding model approximations. Nevertheless, as the number of variables subject to uncertainty increases, encompassing both temporal and cross-variable dependencies in a two-stage optimization program becomes more complex. Furthermore, two-stage programming methods can hardly reflect the dynamic variations of system conditions, especially for multi-stage problems with a sequential structure [13]. As computational complexity is much higher in multi-stage models, there is a trade-off between two- and multi-stage methods for conventional optimization frameworks. In this context, we propose a machine learning (ML) model capable of handling the multi-stage decision-making problem by incorporating the sequential decision process under uncertainty into its learning phase.

In this document, we consider an energy management system (EMS) manag-

ing a PV-ESS plant participating in two different electricity markets: (1) the DA market, where bids encompass energy and capacity for up/down-regulation products for the following day, must be submitted daily and many hours before, and have an hourly granularity; and (2) the RT market, where bids encompass energy products for the following hour, must be submitted hourly and one hour before, and have a 15-min granularity. Thus, bidding decisions for each market must be made with different frequencies and lead times, resulting in a multiple timescale problem. Evidence suggests that simultaneously addressing different timescales in dynamic decision-making under uncertainty can avoid time-inconsistent solutions leading to an improper assessment of risk [14, 15]. Moreover, the independent system operator (ISO) calls for the deployment of procured capacity for up/down-regulation at a 1-min resolution. Therefore the plant's operation must be modeled with a finer granularity. An affine control law is obtained from an explicit model predictive control (MPC) reference-tracking formulation to control the PV-ESS injections to the grid. The reference signal to be tracked derives from the EMS's bidding decisions, uncertainty realizations in PV generation, and ISO's requests for up/down-regulation deployment.

Recently, ML techniques have been used for sequential decision making by combining reinforcement learning (RL) and artificial neural networks (ANNs) [16]. Previous works using RL to make sequential bidding in electricity markets have focused on single-agents either operating on a single market or simultaneously bidding to all markets operating on the same timescales [17, 18, 19]. In theory, a single agent could be trained to bid in multiple markets at different time-scales using multi-task learning [20], saving computation at inference time. Unfortunately, using a single agent could lead to inferior overall performance, as task objectives can compete [21]. Intuitively, our tasks (bidding in the DA and RT markets) are different enough to deteriorate the generalization performance, as each task handles different input information, lead-times, granularity, bidding horizons, and action spaces. Our multi-timescale context can be significantly more challenging than single timescale problems, as the observations of the environment, the rewards, and the actions may have different timescales, resolutions, and lead-times. Multi-agent reinforcement learning (MARL) [22, 23, 24] can address problems with real-world complexity by making agents learn through interactions with the environment based on received reward signals. However, only a few MARL papers have focused on multi-timescale decision problems [25, 26, 27], and none for this specific application, where we deal with multi-dimensional continuous observation and action spaces and the feasibility set of the RT actions depends on the DA actions, adding complexity not previously tackled in the literature.

In this document, we propose the use of a multi-agent deep reinforcement learning (MADRL) [28, 29, 30] approach capable of operating on multiple timescales. In the proposed MADRL framework, two agents are trained to make DA and RT bidding decisions sequentially. Figure 1 shows a simplified scheme of the formulation adopted in our research, illustrating the interaction between two agents based on multi-view ANNs with recurrent layers (MVANNs) and the environment. Two MVANN-based agents make multi-timescale decisions on a daily and hourly basis (N_1 and N_2) with different lead times (k_1 and k_2) based on observations of the environment's state. Both agents are encouraged to collaborate by a shared reward signal, as usual in a

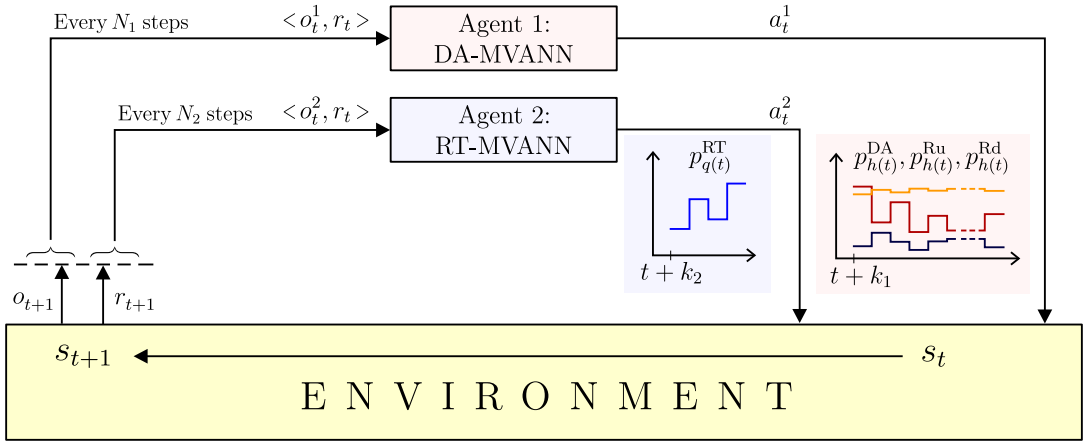


Figure 1: Simplified scheme: two agents interacting with their environment on different timescales

fully cooperative setting [29].

Buffer-rolls are introduced as an experience replay mechanism, containing sufficient information to simulate the PV-ESS controlled operation and the MVANNs' sequential decision-making during a time window, and also capable of adapting to MVANNs' current policies over the learning phase by updating its content. To achieve a cooperative behavior between both MVANN-based agents, we propose a shared reward function that depends on the decisions made by the MVANNs within a mini-batch of buffer-rolls, from which we derive weight updates for both MVANNs. In order to avoid over-exploitation of PV-ESS's resources at the time windows' end, gradient derivation is done with respect to MVANNs' outputs only on user-defined time-steps. To ensure the robustness and reliability of the proposed MADRL framework, we keep track of the adopted bidding policies performance during training using a separate validation set. To compare the performance of the proposed method, we implemented robust and stochastic scenario-based two-stage optimization methods.

Thus, this document aids to fill a gap in the field of ML applications for dynamic decision-making under uncertainty in electric power systems. This research's main contribution is the introduction of a novel MADRL framework to derive efficient energy and AS bids for a hybrid power plant participating on electricity markets operating at different timescales. Furthermore, we propose a innovative approach to solve a multi-timescale multi-agent sequential decision-making problem, where two MVANN-based agents act cooperatively on multi-dimensional continuous observation and action spaces, and where the feasibility set of the second agent's actions depends on the actions of the first one.

3 Background

3.1 Market structure

Power grids coordinate a diverse set of energy assets (e.g., generators, loads, and storage devices) to match supply and demand at all times. Wholesale electricity markets, including those operated by CAISO, PJM, MISO, ISO New England, and New York ISO, follow a two-settlement system in which a DA market seeks to commit transactions based on expected system performance. In contrast, the RT market allows for corrections when the system deviates from expected behavior due to forecast errors or contingencies [31]. Market settlements set prices for multiple products and at different times. Locational marginal prices (LMPs) reflect the marginal value of serving an additional unit of energy at a specified node in the transmission system, typically in (`$/MWh`). Meanwhile, ancillary service marginal prices (ASMPs) compensate AS awards.

We consider a pay-as-cleared and bid-based auction structure in the electricity market. Market participation is limited to self-schedule bids for energy and AS products, which implies quantity-only bids that the ISO will entirely accept [32]. By submitting self-schedule bids, the market participant expresses his willingness to generate/consume at the pointed quantities regardless of the resulting market prices. Due to the relatively small size of the plant, we adopt a price-taker assumption as in [9, 17, 19, 33]. That is, the bidding behavior of the plant has no capability of altering the market-clearing prices as in strategic bidding contexts [34].

We consider that the market participant can submit bids for energy and AS products with an hourly granularity in the DA market and bids for energy products with a 15-min granularity in the RT market. Energy products are expected to be delivered at constant power during the awarded period. We consider the AS products of capacity for up-regulation and down-regulation. The requirement to deploy AS accepted capacity through time is communicated to market participants by a signal sent by the ISO with a 1-min resolution, ranging between zero and the respective awarded capacity. Thus, the market participant must increase/decrease its power injection in response to an up/down-regulation deployment signal. To avoid imbalances between actual and programmed generation through time, the market participant must follow a reference signal according to the time-correspondent awarded DA and RT energy products and the ISO's deployment signal for up/down-regulation. Consequently, the reference signal p_t^{r} used by the control scheme of the plant (see Section 3.2 and 3.3), is a result of the adopted bidding policies from both MVANN-based agents, uncertainty realizations of PV generation and ISO's regulation deployment request.

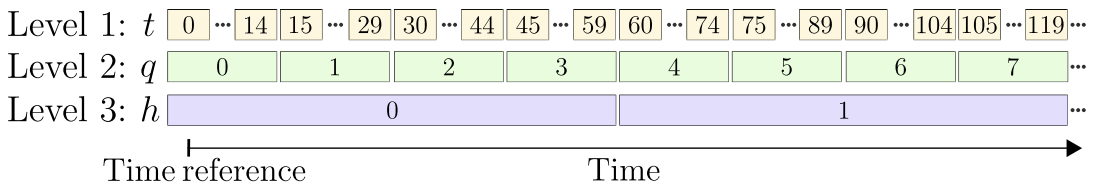


Figure 2: Time discretization representation

A suitable time representation is required, as sequential decisions are made for different time-intervals. For this end, a natural representation for time evolution is employed at three time-interval levels, as illustrated in Fig. 2. Bottom to top, level 3 comprises hourly time-intervals h, level 2 comprises 15-min time-intervals q, and level 1 comprises 1-min time-intervals t. As depicted in Fig. 2, we employ a time reference to set zero values at each level. To transform power units to energy units at different time-intervals, the conversion factors Δ^1, Δ^{15} and Δ^{60} are employed, relatives to the duration of each time-interval with respect to an hour:

$$\Delta^1 = \frac{1}{60}\,\mathtt{h}\ ,\ \ \Delta^{15} = \frac{1}{4}\,\mathtt{h}\ ,\ \ \Delta^{60} = 1\,\mathtt{h}. \tag{1}$$

To ease notation, elements at lower levels are considered contained in elements at higher levels. Equation (2) describes this relationship making use of the floor function $\lfloor x \rfloor$ which gives the largest integer less than or equal to x:

$$q(t) = \left\lfloor t\frac{\Delta^1}{\Delta^{15}} \right\rfloor\ ,\ \ h(t) = \left\lfloor t\frac{\Delta^1}{\Delta^{60}} \right\rfloor. \tag{2}$$

Additionally, a 24-hour time format followed by a day index on parenthesis is used to refer to a particular moment in time, e.g., 13:57 (d), where d is set to zero for the first day that data is available.

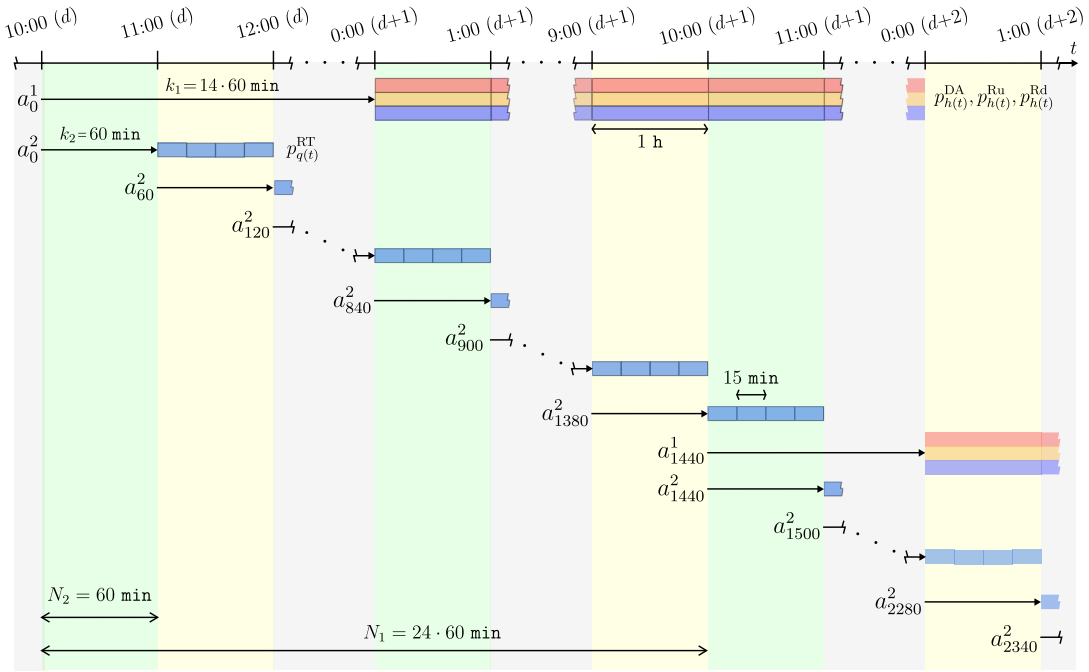


Figure 3: Multi-agent multi-timescale control sequence

Figure 3 illustrates the market program considered in this research using the time representation described earlier. Bids for energy and AS products in the DA market must be submitted by 10 a.m. with hourly granularity covering the 24 hour-intervals of the following day. The DA market establishes schedules for energy and capacity

for regulation and sets related LMPs and ASMPs. These results are published no later than 1 p.m. on the same day of bids submission. Meantime, in the RT market, bids for energy must be submitted one hour before the start of each trading hour and have 15-min granularity. The RT market establishes binding schedules and LMPs for the four 15-min intervals at each hour. These results are published no later than the ending of each trading hour.

3.2 PV-ESS system operation

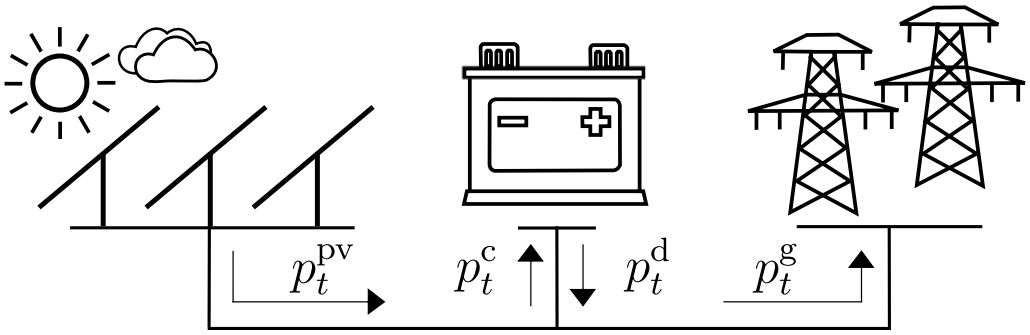


Figure 4: Power balance at the PV-ESS plant connected to the grid

Alongside the problem of submitting self-scheduling bids to the DA and RT markets, this research deals with the controlled operation of the PV-ESS system at a 1-min resolution, which is raised as the following MPC [35] reference-tracking problem:

$$\min_{p_t^{\mathrm{c}}, p_t^{\mathrm{d}}} \ \delta_t^+ + \delta_t^- \tag{3a}$$

$$\text{s.t.} \quad p_t^{\mathrm{r}} = p_t^{\mathrm{g}} + \delta_t^+ - \delta_t^- \tag{3b}$$

$$p_t^{\mathrm{g}} = p_t^{\mathrm{pv}} + p_t^{\mathrm{d}} - p_t^{\mathrm{c}} \tag{3c}$$

$$\check{e}^{\mathrm{s}} \leq e_{t-1}^{\mathrm{s}} + \left(\eta^{\mathrm{c}} p_t^{\mathrm{c}} - \frac{p_t^{\mathrm{d}}}{\eta^{\mathrm{d}}}\right)\Delta^1 \leq \hat{e}^{\mathrm{s}} \tag{3d}$$

$$0 \leq p_t^{\mathrm{c}} \leq y^{\mathrm{s}} \hat{p}^{\mathrm{s}} \ , \ 0 \leq p^{\mathrm{d}} \leq (1 - y^{\mathrm{s}}) \hat{p}^{\mathrm{s}} \tag{3e}$$

$$y^{\mathrm{s}} \in \{0, 1\}, \quad \delta_t^+, \delta_t^- \geq 0 \tag{3f}$$

which aims to minimize the absolute difference $|\delta_t|$ between the reference signal p_t^{r} and the power flow from the hybrid power plant to the power grid p_t^{g}, as (3a) and (3b) depict. This absolute difference is the power imbalance between actual and requested generation at a 1-min scale and is decomposed into under and over-generation as δ_t^+ and δ_t^-, respectively. Equation (3c) models the power balance at the PV-ESS plant connected to the grid, as reflected in Fig. 4. Equations (3d) and (3e) are concerned with the ESS dynamics, while (3f) defines the domain of the variables. This explicit MPC problem is solved as a function of the variables $p_t^* = p_t^{\mathrm{r}} - p_t^{\mathrm{pv}}$ and e_{t-1}^{s}, giving rise to the following affine control law:

$$p_t^{\mathrm{c}}\left(p_t^*, e_{t-1}^{\mathrm{s}}\right) = \mathbb{1}\left(p_t^* \leq 0\right) \min \left\{ \begin{matrix} \textcircled{1} \\ -p_t^* \end{matrix}, \frac{\overset{\textcircled{2}}{\hat{e}^{\mathrm{s}} - e_{t-1}^{\mathrm{s}}}}{\eta^{\mathrm{c}} \Delta^1}, \begin{matrix} \textcircled{3} \\ \hat{p}^{\mathrm{s}} \end{matrix} \right\} \tag{4a}$$

$$p_t^{\mathrm{d}}\left(p_t^*, e_{t-1}^{\mathrm{s}}\right) = \mathbb{1}\left(p_t^* > 0\right) \min \left\{ \begin{matrix} \textcircled{4} \\ p_t^* \end{matrix}, \frac{\overset{\textcircled{5}}{\eta^{\mathrm{d}}(e_{t-1}^{\mathrm{s}} - \check{e}^{\mathrm{s}})}}{\Delta^1}, \begin{matrix} \textcircled{6} \\ \hat{p}^{s} \end{matrix} \right\} \tag{4b}$$

where $\mathbb{1}$ is the indicator function. This mapping comprises piece-wise functions used to drive the operation of the ESS immersed in the hybrid power plant. Following the affine control law derived from MPC's parametric optimization ensures the hybrid plant's physical constraints. Figure 5 illustrates the polyhedral partition sets and corresponding MPC's objective values, where circled numbers denote the polyhedral partitions correspondence to (4). The effectiveness of the bidding models partly relies on an accurate representation of the PV-ESS plant's operation. The affine control law guiding the ESS must be considered by the bidding models to properly simulate the minute-by-minute operation of the hybrid power plant. Equation (4) and Fig. 5a show that a piece-wise formulation is adequate and accurate for the control model representation. Figure 5b shows the objective function (3a) convexity and performance, evidencing a perfect reference tracking for zones ① and ④.

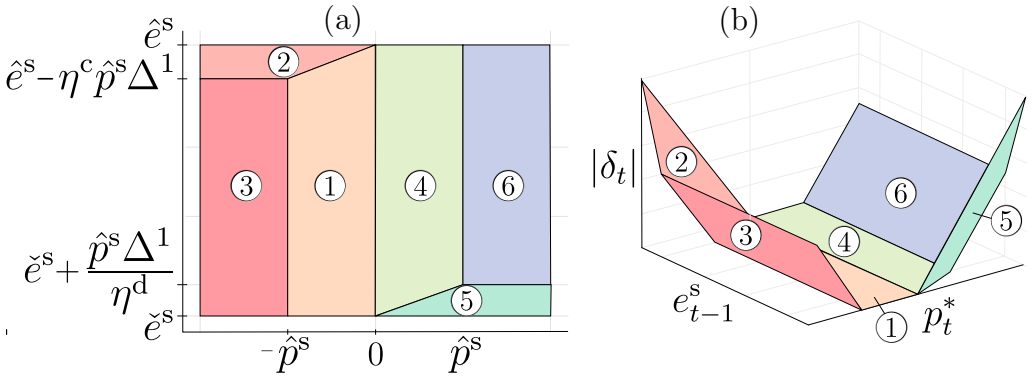


Figure 5: Visual representation of parametric optimization results: (a) Affine control law polyhedral sets (b) Explicit MPC's objective value

3.3 PV-ESS efficient multi-timescale bidding in the DA and RT markets

The efficient multi-timescale bidding of a hybrid power plant is achieved by effectively scheduling and allocating plant resources in the DA and RT markets throughout time and properly delivering awarded products to the power grid under uncertainty. Notice that the EMS faces price, production, and ISO's regulation signal uncertainties (similarly to what would happen in real-world applications), as it does not know in advance the realizations of the uncertain variables. By this means the EMS aims to maximize the following stochastic joint-optimization problem:

$$\max \sum_{t\in\mathcal{T}} \Big(\tilde{\lambda}^{\mathrm{DA}}_{h(t)} p^{\mathrm{DA}}_{h(t)} + \tilde{\lambda}^{\mathrm{Ru}}_{h(t)} p^{\mathrm{Ru}}_{h(t)} + \tilde{\lambda}^{\mathrm{Rd}}_{h(t)} p^{\mathrm{Rd}}_{h(t)} + \tilde{\lambda}^{\mathrm{RT}}_{q(t)} p^{\mathrm{RT}}_{q(t)} - \lambda^{\mathrm{imb}} \left| p^{\mathrm{r}}_t - \tilde{p}^{\mathrm{pv}}_t - p^{\mathrm{d}}_t + p^{\mathrm{c}}_t \right| \Big) \Delta^1 \tag{5a}$$

$$\text{s.t. } p^{\mathrm{r}}_t = p^{\mathrm{DA}}_{h(t)} + \tilde{b}^{+}_t p^{\mathrm{Ru}}_{h(t)} - \tilde{b}^{-}_t p^{\mathrm{Rd}}_{h(t)} + p^{\mathrm{RT}}_{q(t)} \qquad \forall t \in \mathcal{T} \tag{5b}$$

$$e^{\mathrm{s}}_t = e^{\mathrm{s}}_{t-1} + \left(\eta^{\mathrm{c}} p^{\mathrm{c}}_t - \frac{p^{\mathrm{d}}_t}{\eta^{\mathrm{d}}} \right) \Delta^1 \qquad \forall t \in \mathcal{T} \tag{5c}$$

$$\left(p^{\mathrm{DA}}_{h(t)}, p^{\mathrm{Ru}}_{h(t)}, p^{\mathrm{Rd}}_{h(t)}, p^{\mathrm{RT}}_{q(t)} \right) \in \Pi^{\mathrm{m}}_t \qquad \forall t \in \mathcal{T}' \tag{5d}$$

$$\left(p^{\mathrm{DA}}_{h(t)}, p^{\mathrm{Ru}}_{h(t)}, p^{\mathrm{Rd}}_{h(t)}, p^{\mathrm{RT}}_{q(t)} \right) \in \Pi^{\mathrm{b}}_t \qquad \forall t \in \mathcal{T} \tag{5e}$$

$$(4a), (4b) \qquad \forall t \in \mathcal{T}$$

where the objective function (5a) maximizes the profit considering the DA and RT markets incomes for each specific product over the time-intervals included in the optimization horizon $t \in \mathcal{T}$. To keep the hybrid power plant power generation close to the ISO's request, we incorporate an imbalance regularization mechanism to settle deviations between actual and requested generation with a 1-min resolution at a penalization value λ^{imb}. Since the power plant's remunerations depend on the market design, an upper bound for imbalance pricing is chosen in this work, i.e., a high price for the imbalances. Parameters that can be subject to uncertainty are noted with the symbol $\tilde{x}$. Notice that certain price related parameters can be known beforehand for some future periods, according to the market rules detailed in Section 3.1. To keep track of sequentially self-scheduled products in both markets, the control reference signal p^{r}_t is derived in accordance to the awarded energy products and ISO's regulation signal at respective 1-min intervals in (5b). The ISO's requirement to deploy AS accepted capacity for up-regulation and down-regulation are constructed using b^{+}_t and b^{-}_t, whose values correspond, respectively, to the positive and negative parts of a signal b_t (ranging between -1 and 1). Equations (4a), (4b), (5b), and (5c) are used to simulate the hybrid power plant operation under uncertainty in accordance to the affine control law discussed in Section 3.2. Set Π^{m}_t in (5d) fixes decision variables for time-intervals $t \in \mathcal{T}'$ related to self-scheduled products that have been previously submitted to the ISO at the time of solving this problem, in accordance to Section 3.1. Set Π^{b}_t imposes domain restrictions to submit reasonable self-schedule bids to markets and avoid degenerate solutions.

$$\begin{aligned} \Pi^{b}_t = \Big\{ \Big(p^{\mathrm{Rup}}_{h(t)}, p^{\mathrm{Rdn}}_{h(t)}, p^{\mathrm{DA}}_{h(t)}, p^{\mathrm{RT}}_{q(t)} \Big) \in \mathbb{R}^4 \ : \\ 0 \le p^{\mathrm{Ru}}_{h(t)} \le \hat{\beta} \ , \ 0 \le p^{\mathrm{Rd}}_{h(t)} \le \hat{\gamma} \ , \\ \check{\alpha} \le p^{\mathrm{DA}}_{h(t)} \le \hat{\alpha} \ , \text{ and } \check{\alpha} \le p^{\mathrm{RT}}_{q(t)} + p^{\mathrm{DA}}_{h(t)} \le \hat{\alpha} \Big\}. \end{aligned} \tag{6}$$

A multi-stage stochastic optimization method can be used to approximate this problem, where the EMS in charge of the hybrid power plant can derive its self-schedule bids in the DA and RT markets through a two-stage procedure. In the first stage, the EMS determines its bidding in the DA market for hourly energy and

AS products at 10 a.m. each day. Meanwhile, the second stage comprises the self-scheduling of 15-min energy products in the RT market for the hour-ahead period each hour. Since the EMS has to make decisions in the RT market for periods with already procured DA products, decisions made in the first stage affect the decisions in the second stage.

The performance of problem (5) depends on the representation of the system dynamics, controller design, and uncertainty modeling. This problem has been previously approximated by a two-stage formulation for different energy systems, such as in [9] and [4]. In this context, the use of MVANN-based agents to make bidding decisions in both markets can turn beneficial because of MVANN's state-of-the-art ability to fit complex maps and handle uncertainty in time series [36]. When enough data is available, MVANNs can be more computationally efficient for modeling complex problems than conventional optimization approaches [37]. We propose a MADRL framework with a learning phase driven to maximize (5a) by using a shared reward function that depends on the adopted agents' policies in a simulated environment. The environment is simulated by employing historical data, and ensuring that the physical and financial constraints of the optimization problem in (5) are met. This is achieved by employing the MPC affine control law (4a)-(4b) and handling MVANNs' actions in accordance to the market structure. Appendix B presents the detailed model formulation for both scenario-based stochastic and robust optimization stages.

4 Methodology

The proposed methodology considers the formulation, implementation, and test of a novel MADRL framework, comparing it against scenario-based optimization frameworks with stochastic and robust objective functions. These different approaches share the goal of maximizing the profit of a PV/ESS power plant over time. The joint-optimization objective function comprises making DA, RT, and AS bidding decisions for energy and capacity for up/down regulation products (specific AS product) meanwhile minimizing energy imbalance through an imbalance pricing mechanism.

Performance testing for the different frameworks considers using open-access databases of PV generation and prices of different electricity market products. This Section intends to introduce the proposed methodology to test the hypothesis and achieve the research objectives in a staggered manner.

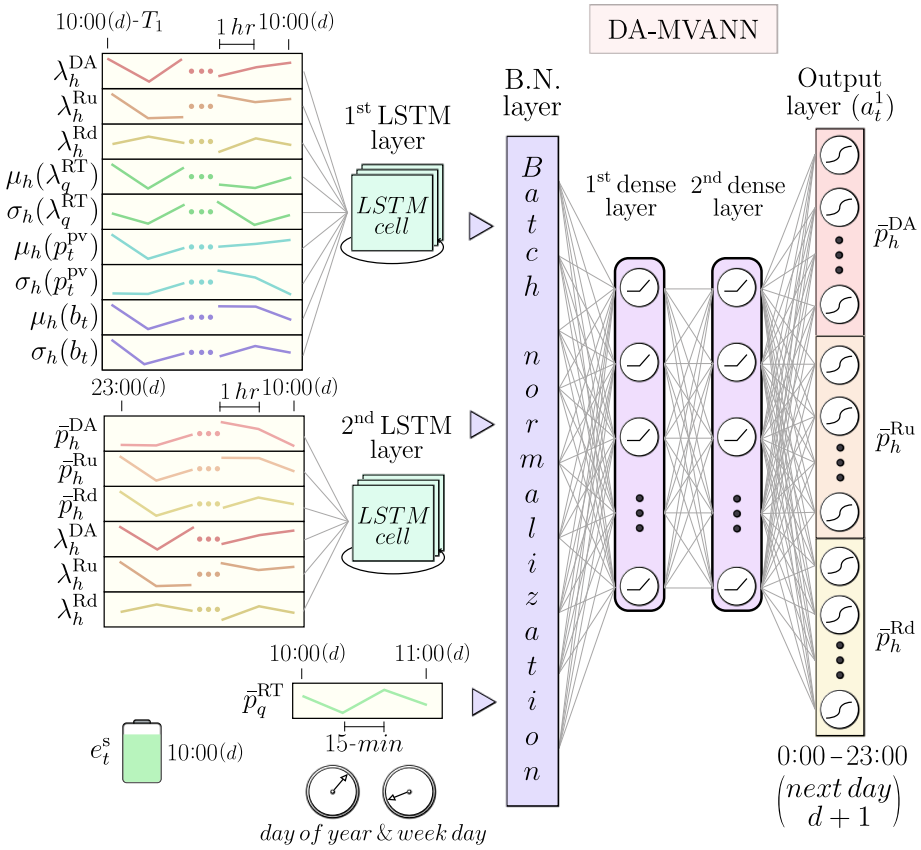
4.1 MADRL for efficient multi-timescale bidding in the DA and RT markets

In order to improve market participation, we introduce an implementation of MADRL for the efficient bidding and operation of a PV-ESS system. ANNs with well-known function approximation properties are employed to learn non-linear mappings to adopt DA and RT cooperative bidding policies as outputs. Two different MVANN-based agents are employed to make bidding decisions for the DA and RT markets, namely DA-MVANN and RT-MVANN as agents 1 and 2, respectively. The inputs of each MVANN-based agent use only currently available information following the market structure discussed in Section 3.1: information related to electricity market products, previously made bidding decisions, PV generation, stored energy, and time representations. For a given time window, we can make bidding decisions from MVANNs following the market program by adjusting agents' timescales (N_1 and N_2) and lead times (k_1 and k_2). On the one hand, financial constraints are ensured by handling MVANNs' outputs, in accordance to (5d) and (5e) (See Section 3.1). On the other hand, physical constraints (i.e. hybrid power plant's injections and storage evolution) are ensured by (5b), (5c), and following the affine control law (4). Initial conditions are required for simulation purposes, as discussed in Section 4.3.

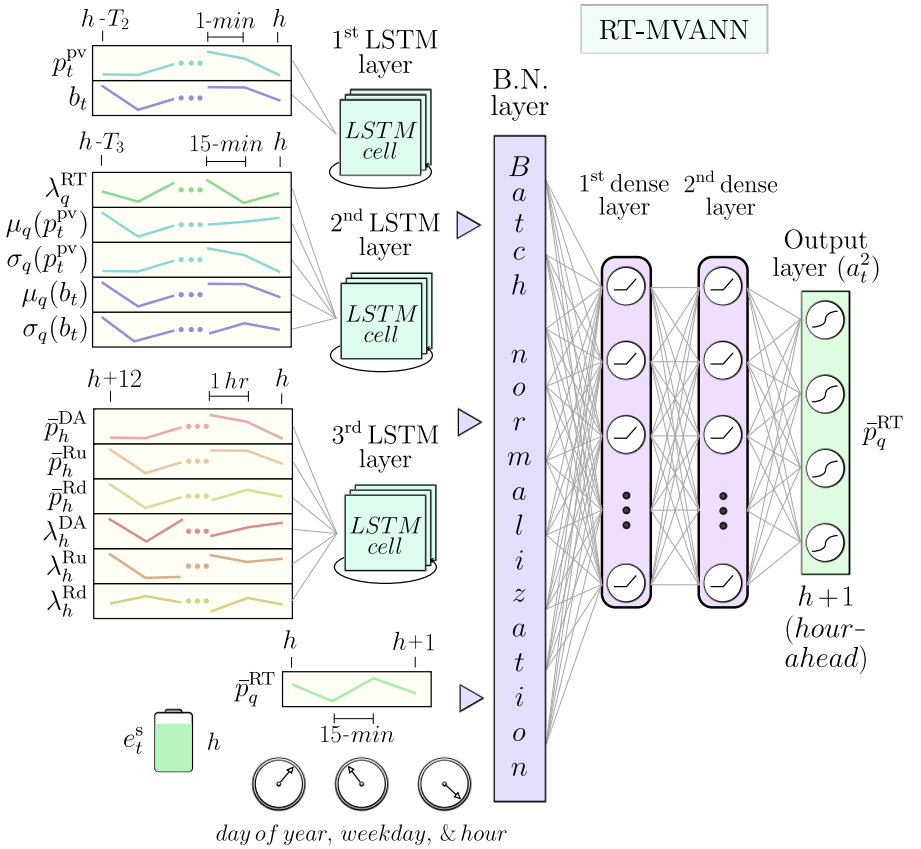
To adjust both MVANN-based agents' policies, we require historical information for market-clearing prices, ISO's reference signal for capacity deployment for up/down-regulation (scaled), and PV generation. Considering the price-taker assumption, we can use historical market clearing prices on the environment's simulation. For practical implementation, in case of lack of historical PV production information, synthetic data could be generated from 1-min irradiance measurements or local weather information, such as in [38].

Note that optimization, bidding submission, and price reception are assumed to be immediate processes in this work, as related timeouts would depend on external

factors. A readjustment of the market participant decision timeline would be needed for real-world applications based on available computational resources and practical experience.



(a) DA-MVANN architecture



(b) RT-MVANN architecture

Figure 6: Diagrams of the DA-MVANN and RT-MVANN architectures

4.2 MVANNs architectures

The DA-MVANN and RT-MVANN architectures are illustrated in Fig. 6. Giving multi-modal measurements to facilitate extracting readily helpful information to increase ML models performance has become a promising topic, covered by the multi-view representation learning field [39]. Acceptance of multi-view representations for MVANNs' inputs allows to include further relevant available information for model performance improvement, such as contemporaneous images of the sky to provide MVANNs with extra information to better forecast future photovoltaic generation [40]. In our implementation, each MVANN takes as inputs multiple features, such as time series at different time resolutions corresponding to past, current, and future periods, currently stored energy, and two-dimensional time representations. To deal with the ANNs model complexity and performance trade-off, we use operators μ_h/μ_q and σ_h/σ_q to down-sample time series to 60-min/15-min intervals using the mean and standard deviation for the respective intervals.

Long short-term memory (LSTM) layers are employed in the MVANNS' architectures due to their ability to capture both long- and short-term patterns, frequent in power systems time-series [41]. Inputs to the second LSTM layer in the DA-MVANN and at the third LSTM layer in the RT-MVANN are related to already made bidding decisions and revealed product prices for future periods. According to the market rules, the DA-MVANN inputs corresponding to future periods (second LSTM layer inputs) are formed considering the DA market information available at 10 a.m. Meanwhile, the RT-MVANN inputs corresponding to future periods (third LSTM layer inputs) consider a range of 12 hours, as DA information is revealed at most at 1 p.m. every day. Time-series containing information for future periods are time-reverted to keep closer information at the end of associated LSTM layers inputs. Outputs from the LSTM layers, bidding decisions made at the RT market for the next four 15-min intervals, currently stored energy, and two-dimensional time representations enter a batch normalization layer and then into a dense feed-forward architecture with *ReLU* activation functions. The amount of data to consider from the past in each LSTM layer input are adjustable hyper-parameters (T_1, T_2, and T_3). Furthermore, the number of LSTM cells for each LSTM layer l (N_L^l), the number of neurons in each dense layer l (N_D^l), and the number of dense layers, are alsoadjustable hyper-parameters for each MVANN.

The DA-MVANN output layer consists of 72 neurons related to bidding decisions for hourly energy and capacity for up/down-regulation products for the following day. Meanwhile, the RT-MVANN output layer consists of 4 neurons related to 15-min energy product bid decisions for the hour-ahead interval. Both output layers dote with tanh() activation functions are MVANN-based agent actions related to each market products for a given time-period denoted by $\bar{x}$, whose output domain remains between -1 and 1. Following (6), bidding decisions must be feasible. To transform MVANN-based agent actions to the respective bidding domains, a scaled min-max normalization function ϕ transforming a variable x from the domain $[\check{c}, \hat{c}]$ to $[-1, 1]$ is inverted, obtaining ϕ^{-1}. Thus, we ensure that DA and RT markets' bidding constraints (6) are satisfied by transforming agents actions with an invertible function.

$$\phi(x, \check{c}, \hat{c}) = \frac{2x - (\hat{c} + \check{c})}{\hat{c} - \check{c}} \tag{7a}$$

$$\phi^{-1}(x, \check{c}, \hat{c}) = \frac{\check{c} + \hat{c} + x(\hat{c} - \check{c})}{2}. \tag{7b}$$

4.3 Buffer-rolls for data management

To split available data in training, validation, and test sets, information is divided into three time-consecutive sets, as usual in time series manipulation for ML. We use the training set to fit both MVANNs' weights, which are updated at each training iteration. The validation set is used to check MVANN-based agents' performance after each training iteration. Once both agents achieve satisfactory results for the validation set, the test set is used to provide an out-of-sample evaluation of the fitted models.

We use training and validation buffer-rolls to manage available data. Each buffer-roll element is associated with a day d and contains enough information to simulate sequential MVANNs' bidding decisions and the PV-ESS's controlled operation for a horizon of H hours since the start time. For each buffer-roll d, we set the start time and time reference to 10:00 (d). Therefore, consecutive buffer rolls contain information shifted 24 hours in time. Each buffer-roll d requires storing entry information, control information, and initial conditions.

Entry information: Time series information for λ_h^{DA}, λ_h^{Ru}, λ_h^{Rd}, λ_q^{RT}, p_t^{pv}, b_t, and two-dimensional time representations to serve as part of the DA-MVANN's inputs for each 10 a.m. time-step into the buffer-roll's horizon H and as part of the RT-MVANN's inputs for each hour. As pre-process, using only the training set partition, we sum to each time series (excluding time representations) its minimum value plus one, apply a log-transformation (except for p_t^{pv}), and then apply ϕ using the resultant minimum and maximum values for each variable.

Control information: Time series information for λ_h^{DA}, λ_h^{Ru}, λ_h^{Rd}, λ_q^{RT}, p_t^{pv} and b_t to contribute to the simulation of PV-ESS controlled operation and obtain markets payments and penalizations based on bidding decisions for the buffer-roll's horizon H. Each control time series remain as originally acquired, as their transformation would distort the environment simulation.

Initial conditions: DA market decisions $\bar{p}_h^{\mathrm{DA}}$, $\bar{p}_h^{\mathrm{Ru}}$ and $\bar{p}_h^{\mathrm{Rd}}$ for hourly intervals from 10:00 (d) until 23:00 (d), RT market decisions $\bar{p}_q^{\mathrm{RT}}$ for 15-min intervals from 10:00 (d) until 10:45 (d), and ESS stored energy at 10:00 (d). Initial conditions can serve as inputs for the MVANNs as well as for simulation. When serving for simulation, each time series is transformed using the ϕ^{-1} function and the domain bounds depicted in (6). The buffer-roll's initial condition values are randomly initialized under a uniform distribution within the respective domain bounds. Note that DA market decisions must be generated before the RT market decisions to ensure (6).

In order to compensate for the MARL non-stationary pathology [29] where agents

face a moving target due to agents' policies evolution through training, we make initial conditions adaptable through an established communication channel between consecutive buffer-rolls. To this end, the hourly horizon H is forced to be more than 24 hours. Hence, consecutive buffer-rolls overlay for some periods. We use this feature to establish a one-way communication channel from buffer-roll d to buffer-roll $d+1$ to update initial conditions iteratively. This update can be done at the end of buffer-roll d time-windows simulation, where DA-MVANN and RT-MVANN's bidding decisions and ESS's stored energy are available for the periods to which the initial conditions of buffer-roll $d+1$ are linked.

To keep all consecutive buffer-rolls communicated and initial conditions adaptable for a comprehensive evaluation, communication is available between the last training and first validation buffer-roll. The maximum number of buffer-rolls that can be built depends on the maximum of MVANNs architecture parameters T_1, T_2, and T_3, the selected horizon H, and the number of available days on the dataset.

4.4 MVANN-based agents learning phase

As mentioned in Section 4.3, for each buffer-roll d, it is possible to simulate H consecutive hours starting at 10:00 (d). We carry out the learning phase by simulating the MVANN-based agents' bidding decisions and 1-min PV-ESS's controlled operation in a mini-batch of buffer-rolls at each training iteration. Algorithm 1 depicts this learning phase.

Algorithm 1 Learning phase

Require: Initialize MVANNs' weights (θ, ω)
for each training iteration **do**
  Randomly sample a mini-batch of buffer-rolls $\mathcal{D}$
  for h from 0 to $H-1$ **do**
    if h mod $24 = 0$ **then**
      Collect DA bids from DA-MVANN
    end if
    Collect RT bids from RT-MVANN
    for t from $60h$ to $60h+59$ **do**
      $p_t^{\mathrm{r}} = p_{h(t)}^{\mathrm{DA}} + b_t^{+} p_{h(t)}^{\mathrm{Ru}} - b_t^{-} p_{h(t)}^{\mathrm{Rd}} + p_{q(t)}^{\mathrm{RT}}$.
      Execute (4)
      Update e_t^{s} by (5c)
    end for
  end for
  Update buffer-rolls $d \in \mathcal{D}+1$ initial conditions
  Compute R^d for each buffer-roll $d \in \mathcal{D}$ by (8)
  Update (θ, ω) by (9a)-(9b)
  if stop criterion is met **then**
    break
  end if
end for

Following Algorithm 1, the learning phase requires initializing the weights of

both MVANNs, where θ and ω corresponds to DA- and RT-MVANNs' weights, respectively. At each training iteration, we simultaneously carry a simulation for a mini-batch of buffer-rolls. In the simulation, we call the MVANNs to make bidding decisions under the market rules, where DA-MVANN and RT-MVANN outputs are transformed by the ϕ^{-1} function in (7b). Data manipulation is required during the learning phase to build the inputs for both ANNs. For instance, computed bids for an hour-step can be required as MVANNs' inputs for a further hour-step. At each hour, the PV-ESS 1-min controlled operation is simulated by using the power reference p_t^{r}, (4), and (5c). Each buffer-roll $d \in \mathcal{D}$ simulation is done independently of each other.

After simulating H hours, initial conditions are updated for the mini-batch of buffer-rolls $d \in \mathcal{D}+1$ using the communication channel described in Section 4.3. Afterwards, the shared reward function for each buffer-roll d is obtained as follows:

$$r_t^d = \Delta^1\big(\lambda^{\mathrm{DA}}_{h(t)} p^{\mathrm{DA}}_{h(t)} + \lambda^{\mathrm{Ru}}_{h(t)} p^{\mathrm{Ru}}_{h(t)} + \lambda^{\mathrm{Rd}}_{h(t)} p^{\mathrm{Rd}}_{h(t)} + \lambda^{\mathrm{RT}}_{q(t)} p^{\mathrm{RT}}_{q(t)} - \lambda^{\mathrm{imb}} |\delta_t|\big), \tag{8a}$$

$$R^d = \frac{\Delta^1}{H} \sum_{t=0}^{\frac{H}{\Delta^1}-1} r_t^d, \tag{8b}$$

where (8a) comes from the objective function depicted in (5a) for each minute t and (8b) is the average reward signal per minute at each buffer-roll $d \in \mathcal{D}$ for the simulated time-window, functioning as a cumulative reward function over a finite horizon. Note that DA and RT bidding decisions can come from initial conditions or transformed MVANNs' outputs.

In order to update both MVANNs' weights using back-propagation at each training iteration, the gradient of the cumulative shared reward function for a mini-batch of buffer-rolls simulations is calculated with respect to each MVANN's weights for user-defined time-steps. The gradient for each MVANN can be decomposed as follows:

$$\nabla_\theta R^{\mathcal{D}} \approx \frac{1}{|\mathcal{D}||\mathcal{H}^\Theta|} \sum_{d\in\mathcal{D}} \sum_{h\in\mathcal{H}^\Theta} \nabla_{\Theta_h} R^d \nabla_\theta \Theta_h, \tag{9a}$$

$$\nabla_\omega R^{\mathcal{D}} \approx \frac{1}{|\mathcal{D}||\mathcal{H}^\Omega|} \sum_{d\in\mathcal{D}} \sum_{h\in\mathcal{H}^\Omega} \nabla_{\Omega_h} R^d \nabla_\omega \Omega_h, \tag{9b}$$

where Θ_h and Ω_h correspond to DA and RT MVANNs' outputs at hour-step h. The gradient of $R^{\mathcal{D}}$ is derived only with respect to the DA-MVANN and RT-MVANN outputs for user-defined hour-steps contained in the sets $\mathcal{H}^\Theta$ and $\mathcal{H}^\Omega$, respectively. The exclusion of some MVANNs' outputs to compute the gradient relies on the observation that if the gradient of the reward function is taken over MVANNs bidding

decisions for all time-steps, resources at the final time-steps would be fully exploited to maximize rewards (maximize profits), without taking future bidding and operational steps into account. This approach to controlling the MVANNs' behaviors is similarly present on rolling horizon optimization frameworks [42], where an adjustable time horizon and decisions stated as resource variables for user-defined steps are included to regularize resource exploitation.

Note that the MVANNs architectures are independent of each other, as no weights are shared between them. Nevertheless, each MVANN influences the weight update calculation of the other through the shared reward function in (8), looking to achieve a cooperative behavior as both MVANNs share the same goal. As a communication mechanism between agents, specific inputs for each architecture are related to already made bidding decisions, as discussed in Sections 4.2 and 4.3.

At the moment of computing, the gradients depicted in (9), inter-temporal dependencies between variables at different time-steps appear, as we are not relying on a Markov Decision Process assumption [28]. For instance, following (5c) it is possible to discern that variable e_t^s at each buffer-roll propagate inter-temporal dependencies on previously made MVANNs' bidding decisions throughout simulation time-steps, given its relation with (4), and the relation of (4) with (5b). The horizon H must be set with the goal of capturing the impact into the future of bidding decisions made at hour-steps into $\mathcal{H}^{\Theta}$ and $\mathcal{H}^{\Omega}$. In our particular context, we consider that bidding decisions for a given day do not have major effects on decisions to be made a month or week ahead, based on the hybrid power plant's characteristics.

Using mini-batches instead of simulating one buffer-roll d at a time for MVANNs training allows a more accurate representation of the population distribution into the dataset for weights updating. In order to get an exhaustive evaluation at the validation set, validation buffer-rolls must be simulated in the original order of days, running one validation buffer-roll at a time and dismissing the MVANNs' weights update steps depicted in Algorithm 1. Nevertheless, as this evaluation can be computationally expensive, a batch simulation is conducted to obtain an approximate evaluation to meet a stop criterion for the MVANNs learning phase. In the case of the test set evaluation, we simulate all hours consecutively, and the already fitted MVANNs make bidding decisions over time by re-arranging and processing the incoming data for each hour to serve as inputs following how the real-world information flow would be. Our proposed MADRL framework can follow the real-world information flow as MVANNs' architectures were constructed for this end: the test phase considers this information flow. It is unnecessary to calculate reward signals at minute-by-minute PV-ESS simulation in the test set, as its unique purpose is to fit both MVANNs' weights in the learning phase. The proposed learning framework allows the use of state-of-the-art optimizers for MVANNs' weights update, such as Adam, RMSprop, and Adadelta [43].

5 Case study

5.1 Data

Related energy LMPs and capacity for up/down-regulation ASMPs are obtained from California's ISO Open Access Same-time Information System (OASIS) website for the time range from 6/19/2017 0:00 until 6/28/2020 23:00. Please refer to [31] for a better understanding of CAISO's market mechanisms. LMPs are obtained for the generation node TOT210S1_7_N002 located near the border between Inyo County (California) and Las Vegas (Nevada). ASMPs are obtained for the AS_CAISO_EXP region. As stated before, we are using historical market clearing prices on the environment's simulation due to the price-taker assumption. As regulation signals are not publicly available at CAISO sites, PJM's traditional regulation signal is used in its stead. PJM's site [44] counts with a 2 seconds resolution historical database for the years 2018, 2019, and 2020. This signal is downsampled to a 1-min resolution for our purposes by taking the respective period's average. We complete missing data for the year 2017 by using the 2018 signal's time series reversed.

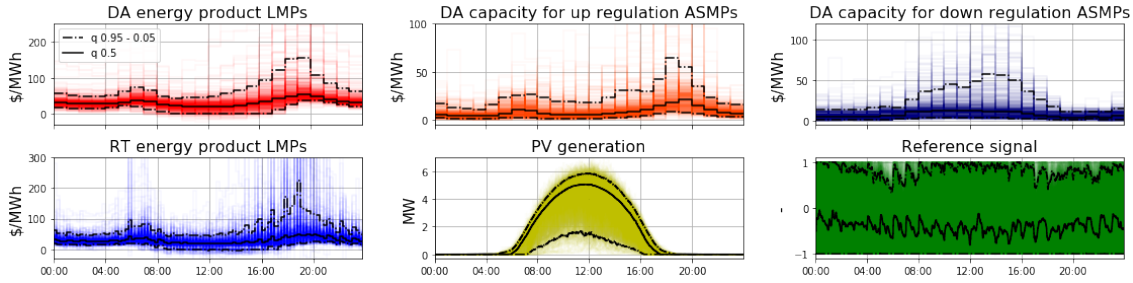


Figure 7: Visualization of dataset from 6/19/2017 0:00 to 6/7/2019 23:00 (First training set).

Regarding PV generation, 1-min resolution data for the time range from 6/19/2017 0:00 to 6/28/2020 23:00 are obtained from an existing PV plant named RTC, NV, Baseline. This power plant is located near Las Vegas and approximately 130 km apart from the CAISO's generation node aforementioned. According to NREL's specifications, this site rate power is 6 MW and features a 30-degree tilt and two strings. This dataset is publicly available at NREL's site [45]. Small missing data time-intervals are handled by performing time-interval averages, while more extensive missing data periods (longer than 30 days) at years 2017 and 2018 are patched with data from 2015 and 2016. Fortunately, for the periods at years 2019 and 2020, only small missing data time-intervals are detected. Figure 7 visualizes this dataset by showing variable values according to the time of the day and respective 5th, 50th, and 95th time-interval quantiles. Because of outliers at the RT energy product LMPs, we clip its values at the top by its quantile 99.9th (training set) only for scenario-generation (for baseline methods) and before pre-processing this time series to serve as entry information for both MVANNs. PV-ESS and market interaction parameters are shown in Table 1.

The dataset is partitioned four times in time-consecutive training, validation, and test sets to evaluate the MADRL framework for a one-year-round period (360

Table 1: PV-ESS and market interaction parameters

$\hat{p}^{\mathrm{s}}$	$\check{e}^{\mathrm{s}}$	$\hat{e}^{\mathrm{s}}$	η_{c}	η_{d}
2.5 MW	0.25 MWh	2.25 MWh	0.9 -	0.95 -
$\check{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	λ^{imb}
0 MWh	6 MWh	2 MWh	2 MWh	200 $/MWh

days). For each of the four partitions, the test set consists of 90 days, and the validation sets correspond to the 30 days before the beginning of each test set. The training sets correspond to all data before the beginning of each validation set. The first partition test set starts at 7/7/2019 00:00.

5.2 MVANNs tuning

Experiments were run on a machine with ST2000DM001-1ER164 disk, Intel (R) Xeon (R) CPU E5-2630 v4 @ 2.2 GHz processor, and an NVIDIA Quadro K620 GPU.

Since the hyper-parameter space is too ample for an exhaustive search, we have performed limited tuning. To select MVANNs hyper-parameters and adjust their weights, a random search of 30 hyper-parameter samples is carried out using both training and validation sets. We have set the horizon simulation length H to 62 hours. $\mathcal{H}^{\Theta}$ is set only for DA-MVANN outputs at 10:00 (d), meanwhile $\mathcal{H}^{\Omega}$ is set for RT-MVANN outputs between 23:00 (d) and 22:00 ($d+1$), included. We have performed weight update calculation using RMSprop on uniformly sampled mini-batches of 16 training buffer-rolls. To adjust the number of training iterations, an early stopping criterion with 50 iterations patience is used [46], which keeps track of the validation buffer-rolls batch cumulative reward function. Weights are initialized using Glorot uniform's initialization. The execution time of each training iteration is approximately 35.9 seconds.

Table 2 shows the selected hyper-parameters for each dataset partition with their respective search spaces, and Fig. 8 shows the training (validation) cumulative mini-batch (batch) rewards for selected MVANNs by dataset partition. We can see in Fig. 8 that for the fourth set partition, the validation cumulative batch reward separates from the training cumulative mini-batch reward. This separation could indicate that the training and validation sets do not represent similar population distributions for the last partition. An existing concept-drift in electricity prices due to pandemic effects and more significant PV generation variability due to winter effects in the fourth partition can partly explain this behavior.

Table 2: Hyper-parameter values for each dataset partition.

MVANN	Hyper-param.	Dataset partition				Search space
		1	2	3	4	
DA-MVANN	T_1	168	168	96	96	72, 96, 120, 144, 168, 192
	N_L^1	18	45	25	21	10-64
	N_L^2	20	32	14	63	10-64
	N_D^1	93	84	88	44	40-100
	N_D^2	-	48	-	22	20-100
RT-MVANN	T_2	3	1	3	1	1-3
	T_3	48	48	24	48	24, 48, 72
	N_L^1	52	60	59	58	10-64
	N_L^2	22	15	20	40	10-64
	N_L^3	44	16	18	20	10-64
	N_D^1	93	92	84	53	40-100
	N_D^2	84	21	-	91	20-100

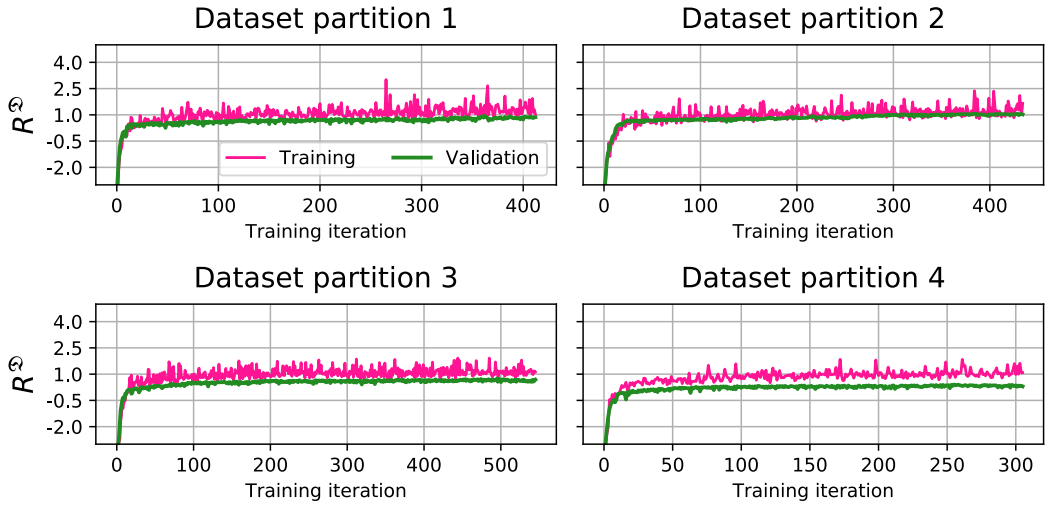


Figure 8: Training and validation shared cumulative rewards ($R^{\mathcal{D}}$) versus number of training iterations for selected MVANNs by dataset partition.

5.3 Scenario-based robust and stochastic optimization

Scenario-based robust (worst-case) and stochastic optimization methods are employed as baselines for the surrogate two-stage optimization problem stated in (5). For the first and second stage formulation (bidding in the DA and RT market), a horizon $|\mathcal{T}|$ of 62 and 12 hours are selected, respectively. Decision variables are stated as recourse variables, with the exemption of the ones required to be submitted to the ISO when solving the problem at each optimization stage, i.e., each variable has an additional dimension s to indicate the scenario to which it is linked. To introduce the presented PV-ESS 1-min explicit control solution stated in (4), it is necessary to include binary variables at each step t and scenario s, dramatically increasing the computational efforts. To keep computational tractability, binary variables are only included for the first two-hour intervals in the second stage of the model and avoided in the first stage.

5.3.1 Scenario generation

An adaptation of k-nearest neighborhood to historical data paths is employed as a scenario-generation technique for both stages, similar to the work done in [47]. This method implies constructing a ranking based on the L^2 norm values of vectors, where these vectors consist of the difference between the last T variable measurements and same-length paths created from historical data of the same variable acquired at equivalent time-intervals in previous days. Once this ranking has been constructed, we select the N_s vectors with lower L^2 norm values, then the next $|\mathcal{T}|$ measurements that follow the ending of each correspondent path serve as scenarios, where $|\mathcal{T}|$ corresponds to the horizon length of each stage. The last T measurements to be considered to generate each variable scenarios are independent for each stage and hour of the day, as different horizon requirements and variable's nature calls for different adjustments. For assessing the quality of generated scenarios for each variable, stage, and hour of the day, we employed the energy score [48] metric ES_T. In the case of equally likely scenarios, the formula is simplified as:

$$ES_T = \frac{1}{|\mathcal{T}|}\sum_{t=1}^{|\mathcal{T}|}\Big(\frac{1}{N_s}\sum_{i=1}^{N_s}|\tilde{\lambda}_t^i - \lambda_t| - \frac{1}{2N_s^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s}|\tilde{\lambda}_t^i - \tilde{\lambda}_t^j|\Big) \tag{10}$$

where $\tilde{\lambda}_t^i$ is the variable value for the scenario i and time-interval t and λ_t is the real variable value at time-interval t. To adjust the length T of the vectors used to select N_s historical data paths of length $|\mathcal{T}|$ to serve as scenarios for each variable, hour of the day, and stage, scenarios are generated for each hour-step of the validation set and related ES's are calculated. We made a grid search with a T hourly equivalent range running from 1 to 193 hours. The number of scenarios N_s is set to 10 for both stages.

Unlike the stochastic and robust formulations, which require scenarios to explicitly represent the uncertainty, our proposed MADRL framework directly maps information that would be available in accordance to the real-world environment under the assumptions. Thus, the MADRL framework does not require an intermediate

step to explicitly represent the uncertainty by forecasting or scenario generation. In other words, the MVANNs learn this mapping during the learning phase driven directly by the finite cumulative reward function stated in (8b). By this means, it adopts an implicit uncertainty representation within the hidden layers.

5.4 Results

This section analyzes the main experimental results by comparing the proposed MADRL framework to scenario-based two-stage robust and stochastic optimization baselines. For a one-year-round period (360 days), the computing time for our proposed method was 0.02 s per hour-step on average, i.e., the time for computing the corresponding bidding decision for each hour-step. Meanwhile, the optimization times for the robust and stochastic methods were 10.63 s and 2.83 s per hour-step on average, respectively. Thus, the computing times of the proposed framework were only 0.18% and 0.7% of the robust and stochastic methods' optimization times.

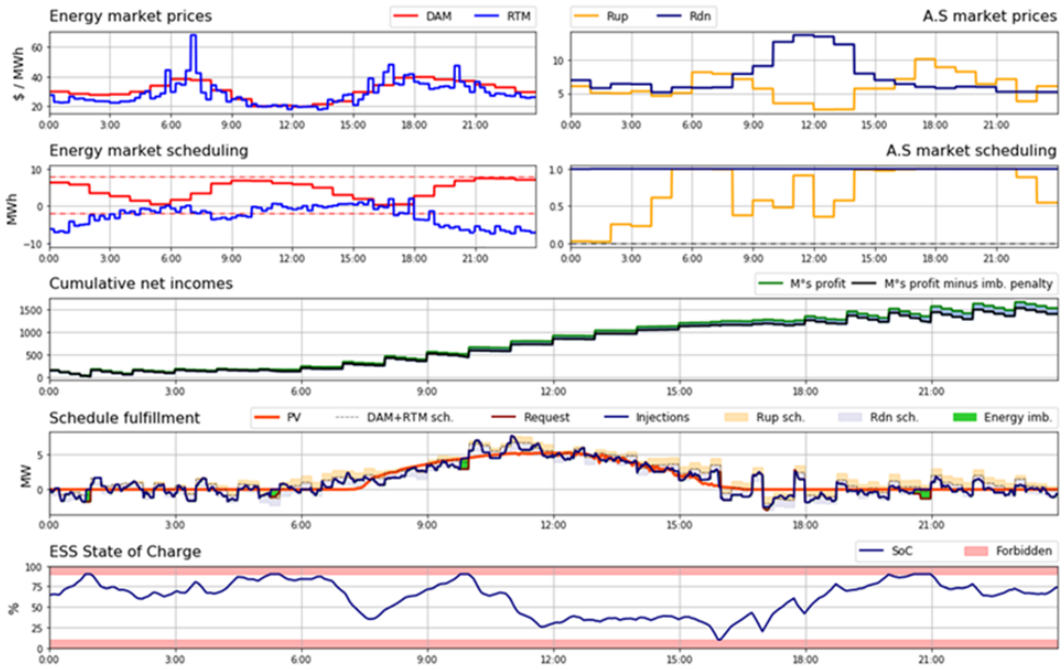


Figure 9: Hybrid PV-ESS plant operation and results for a given test set day

Figure 9 shows results of the plant's operation for a given day of a test set, which is driven by the proposed MADRL agents' adjusted policies for bidding in both electricity markets. From top to bottom row: (i) LMPs and ASMPs for energy and AS products; (ii) DA-ANN and RT-ANN agents' biddings for each electricity product; (iii) cumulative revenues compounded by all electricity products incomes in green and cumulative profits in black at 1-min granularity; (iv) schedule fulfillment shows PV generation, awarded DA and RT energy and AS products, requested reference signal in accordance to MADRL self-scheduled products and ISO's regulation signal, and energy imbalances indicating the difference between agreed and actual generation; and (v) ESS' energy storage evolution through time.

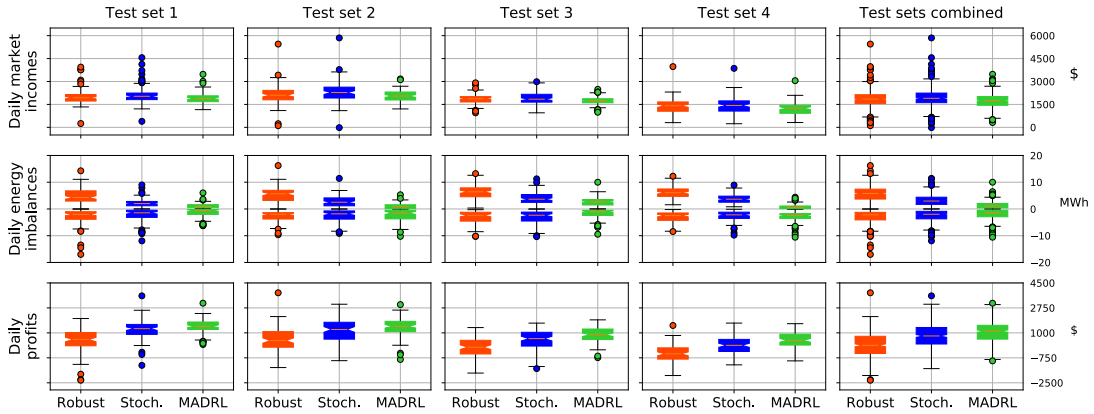


Figure 10: Daily market incomes, energy imbalances (positive domain: over-generation — negative domain: under-generation), and profits boxplots per method and for each dataset partition.

Figure 10 shows boxplots for the daily market incomes, energy imbalances, and profits for the different test sets and methods. While the daily market incomes consist of total payments received by the PV-ESS plant for its participation in the DA and RT markets, the daily energy imbalances consist of the daily sum of over and under-generation at each minute, i.e., δ_τ^- and δ_τ^+. The daily profits comprises both the incomes for each product and the penalizations using the imbalance price λ^{imb}. This factor also serves as a regularization mechanism in (8a) to drive the MVANNs weights' updating for the MADRL approach and in the objective function (5a) for the robust and stochastic methods.

F-tests and t-tests were performed to assess the statistical significance of the differences in variance and mean, respectively, between the results obtained by MADRL and stochastic and robust methods. Table 3 shows statistics for daily market incomes, imbalance penalizations, and profits. The daily market incomes obtained with the proposed MADRL framework have smaller variance than the stochastic and robust implementations (F-test's p-values of 5.66×10^{-9} and 6.45×10^{-5}), while it shows smaller incomes (t-test's p-values of 1.23×10^{-19} and 1.97×10^{-7}).

The daily energy imbalances show statistically significant smaller variance at over and under-generation for the MVANNs implementation when compared against stochastic and robust approaches (F-test's p-values of 7.24×10^{-12} and 1.64×10^{-32} for over-generation, and 2.78×10^{-3} and 1.2×10^{-3} for under-generation). Even more, the MVANNs method achieved smaller values for both energy imbalances, with a bias towards under-generation (t-test's p-values of 2.48×10^{-119} and 7.27×10^{-66} for over-generation, and 4.7×10^{-17} and 1.99×10^{-13} for under-generation). Reference-tracking performance consists of the percentage of 1-min steps where the PV-ESS's power injections to the grid p_τ^{g} matched the power reference signal p_τ^{r}, i.e. $|\delta| = 0$. Results show that while the sum of daily market incomes seems favorable for the baseline methods, they are subject to higher and more dispersed energy imbalances. This behavior, coupled with a more accurate reference-tracking performance of the signal generated by the MVANN-based agents, shows that our approach achieved less variability at controlling the PV-ESS power plant than the baseline approaches.

Table 3: Statistics for daily market incomes, imbalance penalizations, and profits (test sets combined)

		Robust	Stochastic	MADRL
DA energy product ($)	*Mean*	1,789	2,125	1,898
	Std.	1,059	1,548	740
	Sum	642,260	762,989	681,227
Capacity for up-regulation product ($)	*Mean*	321	347	168
	Std.	143	147	89
	Sum	115,128	124,728	60,214
Capacity for down-regulation product ($)	*Mean*	381	382	347
	Std.	233	235	208
	Sum	136,717	137,044	124,628
RT energy product ($)	*Mean*	-647	-914	-678
	Std.	965	1,421	376
	Sum	-232,120	-328,265	-243,549
Daily market incomes ($)	*Mean*	1,844	1,940	1,734
	Std.	598	662	488
	Sum	661,985	696,496	622,520
Imbalance penalizations ($)	*Mean*	1734	1169	697
	Std.	631	582	477
	Sum	622,512	419,641	250,341
Daily profits ($)	*Mean*	110	771	1,037
	Std.	861	875	684
	Sum	39,473	276,854	372,179

Table 4: Daily under/over-generation and reference tracking performance by method (test sets combined)

		Robust	Stochastic	MADRL
Daily under-generation (MWh)	*Mean*	2.95	2.64	2.04
	Std.	2.39	2.36	2.04
	Sum	1,059.07	947.99	730.60
Daily over-generation (MWh)	*Mean*	5.72	3.20	1.45
	Std.	2.67	2.01	1.40
	Sum	2,053.49	1,150.22	521.11
Reference-tracking performance (%)		73.05	79.13	86.63

Table 4 summarizes under/over-generations shown in Fig. 10 and reference-tracking performance for the one-year-round implementation of each method.

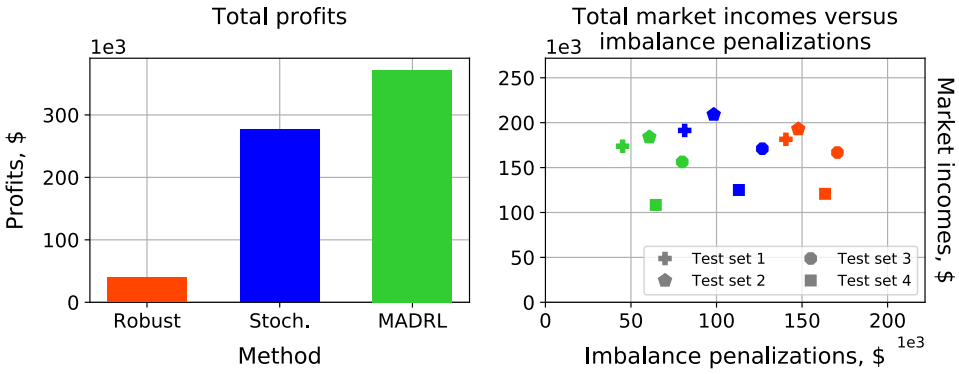


Figure 11: Total market incomes and imbalance penalizations: (a) Total profits per method (b) Total market incomes versus imbalance penalizations per method and for each dataset partition.

Figure 11a shows the total profits for the one-year-round market participation for each method, showing that, from this perspective, our proposed MADRL framework achieved superior performance. The daily profits show statistically significant higher mean (t-test's p-values of 1.44×10^{-80} and 2.98×10^{-19}) and smaller variance (F-test's p-values of 6.53×10^{-6} and 1.73×10^{-6}) than the robust and stochastic methods, respectively. However, Fig. 11b shows that the proposed MADRL framework achieves this higher performance by trading-off market incomes for a better provision of services, i.e., lower energy imbalances. Also, note that all three methods' performances follow a similar trend across test sets, evidencing that all methods captured the effects of winter (first and fourth test sets) and coronavirus (third and fourth test sets).

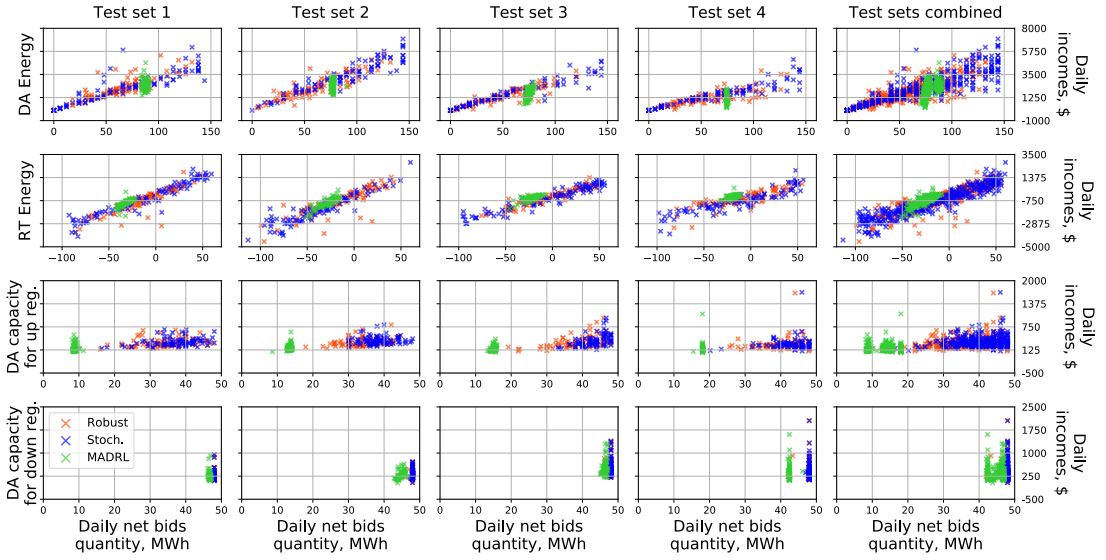


Figure 12: Daily incomes versus daily net bid quantity per method for each market product.

In order to better understand each method's bidding strategy, Fig. 12 shows each market product daily incomes against daily net self-scheduled bids quantity

by method, where the latter refers to the total sum of self-scheduled products for each daily period in (MWh). We observe that the DA-MVANNs and RT-MVANNs derived a different strategy for each dataset partition, as each one concentrates its bids around different values and with different levels of dispersion, where the latter is caused due to each MVANN sensitivity to input values. The proposed MADRL framework achieved statistically significant smaller bids quantity variance for each product, with the exemption of the capacity for down-regulation product. F-test's p-values of 3.39×10^{-41} and 8.33×10^{-12} (DA energy), 4.98×10^{-22} and 4.01×10^{-20} (up-regulation), 8.1×10^{-3} and 1.25×10^{-2} (down-regulation), and 3.46×10^{-111} and 1.09×10^{-62} (RT energy).

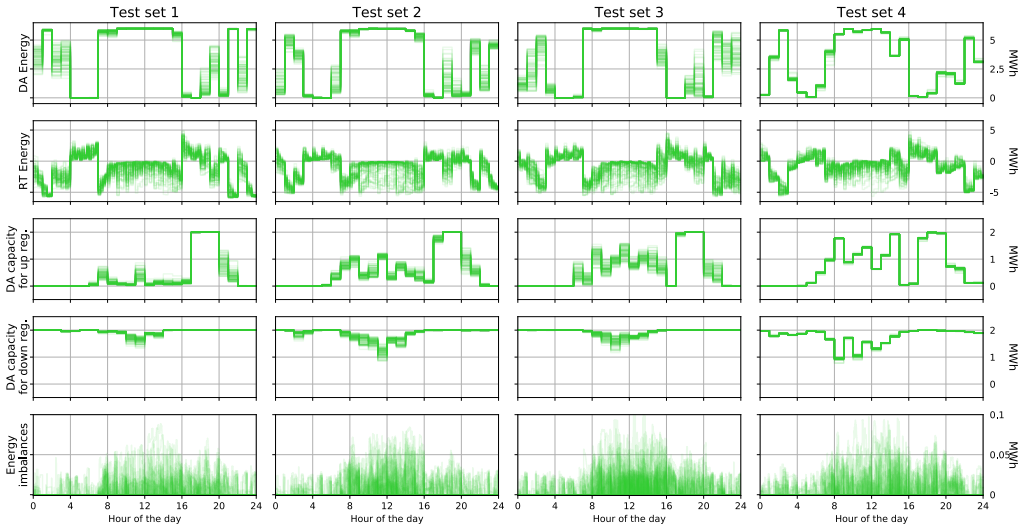


Figure 13: MADRL submitted bis and energy imbalances on each day for each test set.

Figure 13 shows the submitted bids for each product and measured energy imbalances against the time of the day for each test seat. We observe that adjusted MVANNs for each test set adopted different yet similar bidding strategies, such as Fig. 12. Nevertheless, it allows us to appreciate that adopted strategies are robust between days, only existing minor decision-making deviations. Therefore, our proposed MADRL method achieves superior performance by finding a pattern to follow every day, being susceptible to variations in the input in specific day hours for each product. For example, the DA energy product bids seem to be rigorously high at mid-day. Nevertheless, it can be seen as a higher susceptibility to MVANNs inputs at night. In contrast, the RT energy product bidding seems to follow the opposite behavior, presenting higher variations at mid-day.

We ran two additional one-year-round simulations to evaluate how using storage and participating in AS markets may improve the economic viability of investing in ESS. These simulations consisted of 1) the PV-ESS power plant participating only in the energy markets (i.e., without participating in AS markets), and 2) the PV power plant without storage selling all its injections at the RT market energy price. According to our results, cases 1 and 2 would reduce total market incomes by 61.1% and 67.7%, respectively. Therefore, using an ESS and participating in both energy and AS markets would increase the total market incomes by approximately 460 k$

in a year.

6 Conclusions

This work proposed a MADRL framework to derive efficient bidding strategies to allocate energy and AS products in the DA and RT markets operating with different timescales and lead times, while ensuring a feasible physical and financial operation of the PV-ESS hybrid power plant. Furthermore, we introduced a novel approach to solve a multi-timescale multi-agent sequential decision-making problem that achieves competitive results against an implementation of scenario-based two-stage robust and stochastic optimization. Based on the experimental setup, we observe that our MADRL framework shows: (i) higher total profits; (ii) comparable mean values for daily market incomes; (iii) smaller variance for daily market incomes, energy imbalances, and daily net bid quantities; and (iv) better resource allocation based on reference tracking performance and energy imbalance results. In future work, the performance of the proposed method could be further improved by the inclusion of additional information, the use of different architectures, or a more exhaustive hyper-parameter search. Moreover, although outside the scope of our work, a performance comparison of single versus multiple agents could be relevant from a ML perspective.

A key feature of our approach is its flexibility to adapt to new environments from the points of view of market modeling and the control strategy used in a hybrid power plant with storage. For example, due to the price-taker assumption, the MVANN-based agents considers independence of the external variables regarding the EMS's bidding decisions. However, this assumption could be relaxed by implementing a market simulator into the MVANNs learning phase to obtain market-clearing prices at computational and complex modeling expenses. Moreover, the proposed implementation could be adapted to other hybrid power plant control methods, as long as sufficient information is available to simulate its operation.

References

[1] S. R. Sinsel, R. L. Riemke, V. H. Hoffmann, Challenges and solution technologies for the integration of variable renewable energy sources—A review, Renewable Energy 145 (2020) 2271–2285.

[2] F. J. Heredia, M. D. Cuadrado, C. Corchero, On optimal participation in the electricity markets of wind power plants with battery energy storage systems, Computers & Operations Research 96 (2018) 316–329.

[3] M. U. Hashmi, W. Labidi, A. Bušić, S.-E. Elayoubi, T. Chahed, Long-term revenue estimation for battery performing arbitrage and ancillary services, in: 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2018, pp. 1–7.

[4] R. Khatami, K. Oikonomou, M. Parvania, Look-ahead optimal participation of compressed air energy storage in day-ahead and real-time markets, IEEE Trans. Sustain. Energy 11 (2) (2020) 682–692.

[5] A. Shapiro, A. Nemirovski, On complexity of stochastic programming problems, in: Continuous Optimization: Current Trends and Modern Applications, Springer US, Boston, MA, 2005, pp. 111–146.

[6] E. Akbari, R.-A. Hooshmand, M. Gholipour, M. Parastegari, Stochastic programming-based optimal bidding of compressed air energy storage with wind and thermal generation units in energy and reserve markets, Energy 171 (2019) 535–546.

[7] J. Aghaei, M. Barani, M. Shafie-khah, A. A. Sánchez de la Nieta, J. P. S. Catalão, Risk-constrained offering strategy for aggregated hybrid power plant including wind power producer and demand response provider, IEEE Transactions on Sustainable Energy 7 (2) (2016) 513–525.

[8] O. Lak, M. Rastegar, M. Mohammadi, S. Shafiee, H. Zareipour, Risk-constrained stochastic market operation strategies for wind power producers and energy storage systems, Energy 215 (2021) 119092.

[9] M. Rahimiyan, L. Baringo, Strategic bidding for a virtual power plant in the day-ahead and real-time markets: A price-taker robust optimization approach, IEEE Trans. Power Syst. 31 (4) (2016) 2676–2687.

[10] A. Akbari-Dibavar, V. Sohrabi Tabar, S. Ghassem Zadeh, R. Nourollahi, Two-stage robust energy management of a hybrid charging station integrated with the photovoltaic system, International Journal of Hydrogen Energy 46 (24) (2021) 12701–12714.

[11] J. L. Crespo-Vazquez, C. Carrillo, E. Diaz-Dorado, J. A. Martinez-Lorenzo, M. Noor-E-Alam, Evaluation of a data driven stochastic approach to optimize the participation of a wind and storage power plant in day-ahead and reserve markets, Energy 156 (2018) 278–291.

[12] E. Roos, D. den Hertog, Reducing conservatism in robust optimization, INFORMS Journal on Computing 32 (4) (2020) 1109–1127.

[13] Y. C. Han, G. H. Huang, C. H. Li, An interval-parameter multi-stage stochastic chance-constrained mixed integer programming model for inter-basin water resources management systems under uncertainty, in: 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Vol. 5, 2008, pp. 146–153.

[14] B. Rudloff, A. Street, D. M. Valladão, Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences, European Journal of Operational Research 234 (3) (2014) 743–750.

[15] A. Brigatto, A. Street, D. M. Valladao, Assessing the cost of time-inconsistent operation policies in hydrothermal power systems, IEEE Transactions on Power Systems 32 (6) (2017) 4541–4550.

[16] K. Arulkumaran, M. P. Deisenroth, M. Brundage, A. A. Bharath, Deep reinforcement learning: A brief survey, IEEE Signal Process. Mag. 34 (6) (2017) 26–38.

[17] D. Cao, W. Hu, X. Xu, T. Dragičević, Q. Huang, Z. Liu, Z. Chen, F. Blaabjerg, Bidding strategy for trading wind energy and purchasing reserve of wind power producer – A DRL based approach, Int. Journal of Electrical Power & Energy Systems 117 (2020) 105648.

[18] R. Chen, I. C. Paschalidis, M. C. Caramanis, P. Andrianesis, Learning from past bids to participate strategically in day-ahead electricity markets, IEEE Trans. Smart Grid 10 (5) (2019) 5794–5806.

[19] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, G. Strbac, Deep reinforcement learning for strategic bidding in electricity markets, IEEE Trans. Smart Grid 11 (2) (2020) 1343–1355.

[20] Y. Zhang, Q. Yang, A survey on multi-task learning, IEEE Transactions on Knowledge and Data Engineering (2021) 1–1.

[21] T. Standley, A. R. Zamir, D. Chen, L. Guibas, J. Malik, S. Savarese, Which tasks should be learned together in multi-task learning? (2020). arXiv:1905.07553.

[22] R. Lu, Y.-C. Li, Y. Li, J. Jiang, Y. Ding, Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management, Applied Energy 276 (2020) 115473.

[23] L. Xi, J. Chen, Y. Huang, Y. Xu, L. Liu, Y. Zhou, Y. Li, Smart generation control based on multi-agent reinforcement learning with the idea of the time tunnel, Energy 153 (2018) 977–987.

[24] J. Wang, L. Sun, Dynamic holding control to avoid bus bunching: A multi-agent deep reinforcement learning framework, Transportation Research Part C: Emerging Technologies 116 (2020) 102661.

[25] J. Wu, K. Li, Q.-S. Jia, Decentralized multi-agent reinforcement learning with multi-time scale of decision epochs, in: 2020 59th IEEE Conference on Decision and Control (CDC), 2020, pp. 578–584.

[26] J. Shin, J. H. Lee, Multi-timescale, multi-period decision-making model development by combining reinforcement learning and mathematical programming, Computers & Chemical Engineering 121 (2019) 556–573.

[27] C. Wernz, Multi-time-scale Markov decision processes for organizational decision-making, EURO Journal on Decision Processes 1 (3-4) (2013) 299–324.

[28] P. Hernandez-Leal, B. Kartal, M. E. Taylor, A survey and critique of multiagent deep reinforcement learning, Autonomous Agents and Multi-Agent Systems 33 (6) (2019) 750–797.

[29] S. Gronauer, K. Diepold, Multi-agent deep reinforcement learning: a survey, Artificial Intelligence Review (2021).

[30] W. Du, S. Ding, A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications, Artificial Intelligence Review 54 (5SN - 1573-7462) (2021) 3215–3238.

[31] A. W. Dowling, R. Kumar, V. M. Zavala, A multi-scale optimization framework for electricity market participation, Applied Energy 190 (2017) 147–164.

[32] California Independent System Operator (2021). [link].
URL https://www.caiso.com/Documents/Section30_Bid-Self-ScheduleSubmission_CAISOMarkets_asof_Feb15_2018.pdf

[33] J. Hu, M. R. Sarker, J. Wang, F. Wen, W. Liu, Provision of flexible ramping product by battery energy storage in day-ahead energy and reserve markets, IET Generation, Transmission & Distribution 12 (2018) 2256–2264(8).

[34] C. N. Dimitriadis, E. G. Tsimopoulos, M. C. Georgiadis, Strategic bidding of an energy storage agent in a joint energy and reserve market under stochastic generation, Energy (2021) 123026.

[35] F. Borrelli, A. Bemporad, M. Morari, Predictive control for linear and hybrid systems, Cambridge University Press, 2017.

[36] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.

[37] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: A survey, Heliyon 4 (11) (2018).

[38] J. Bright, C. Smith, P. Taylor, R. Crook, Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data, Solar Energy 115 (2015) 229–242.

[39] R. Zhang, F. Nie, X. Li, X. Wei, Feature selection with multi-view data: A survey, Information Fusion 50 (2019) 158–167.

[40] Y. Sun, G. Szűcs, A. R. Brandt, Solar PV output prediction from video streams using convolutional neural networks, Energy Environ. Sci. 11 (2018) 1811–1818.

[41] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8) (1997) 1735–1780.

[42] W. B. Powell, Clearing the jungle of stochastic optimization, in: Bridging Data and Decisions, INFORMS TutORials in Operations Research, 2014, pp. 109–137.

[43] R. Zaheer, H. Shaziya, A study of the optimization algorithms in deep learning, in: 2019 Third Int. Conf. on Inventive Systems and Control (ICISC), 2019, pp. 536–539.

[44] Pennsylvania-New Jersey-Maryland Interconnection (2021). [link]. URL https://www.pjm.com/markets-and-operations/ancillary-services.aspx

[45] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, J. Shelby, The national solar radiation data base (NSRDB), Renewable and Sustainable Energy Reviews 89 (2018) 51–60.

[46] Y. Bengio, Practical recommendations for gradient-based training of deep architectures, in: Neural networks: Tricks of the trade, Springer, 2012, pp. 437–478.

[47] W. J. Raseman, B. Rajagopalan, J. R. Kasprzyk, W. Kleiber, Nearest neighbor time series bootstrap for generating influent water quality scenarios, Stochastic Environmental Research and Risk Assessment 34 (1) (2020) 23–31.

[48] D. Sari, Y. Lee, S. Ryan, D. Woodruff, Statistical metrics for assessing the quality of wind power scenarios for stochastic unit commitment, Wind Energy 19 (5) (2015) 873–893.

Appendix A: Publications

Related to the current work:

- **Ochoa, Tomás**; Gil, Esteban; Angulo, Alejandro; Valle, Carlos. 2022. "Multi-agent Deep Reinforcement Learning for Efficient Multi-Timescale Bidding of a Hybrid Power Plant in Day-Ahead and Real-Time Markets". Accepted in Applied Energy (WoS Impact Factor: 9.746 (2020)), in press.
- **Ochoa, Tomás**; Gil, Esteban; Angulo, Alejandro. 2022. "Efficient Bidding of a PV Power Plant with Energy Storage Participating in Day-Ahead and Real-Time Markets Using Artificial Neural Networks". Accepted for presentation in 2022 IEEE PES General Meeting.

Not related to the current work:

- Gil, Esteban; Morales, Yerel; **Ochoa, Tomás**. 2021. "Addressing the Effects of Climate Change on Modeling Future Hydroelectric Energy Production in Chile" Energies (WoS Impact Factor: 3.004 (2020)) 14, no. 1: 241. https://doi.org/10.3390/en14010241.
- Serpell, Cristian; **Ochoa, Tomás**; Gil, Esteban; Valle, Carlos. 2022. "Power Load Probabilistic Forecasting and Scenario Generation with Mixture Density Networks". To be submitted to IEEE Transactions on Power Systems, currently in writing stage.

Appendix B: Two-stage optimization model formulation

First stage

Parameters and sets

The parameters needed to formulate the first stage of scenario-based optimization models are:

$$q^{\mathrm{RT}} = 4\ ,\ h^{\mathrm{DA}} = 14\ ,\ \mathcal{T}^{\mathrm{DA}} = \left\{1, ..., \frac{48 + h^{\mathrm{DA}}}{\Delta^1}\right\}\ ,\ \Xi = \{1, ..., 10\} \tag{11}$$

where h^{DA} and q^{RT} indicates the time steps where bidding decisions where already made for the DA and RT market, respectively. $\mathcal{T}^{\mathrm{DA}}$ and Ξ are the sets containing optimization's time horizon in 1-min resolution and generated scenarios indexes, respectively.

Decision variables

The decision variables considered in the first stage of scenario-based optimization models are:

$$p^{\mathrm{g}}_{t,\xi}, p^{\mathrm{d}}_{t,\xi}, p^{\mathrm{c}}_{t,\xi}, p^{\mathrm{r}}_{t,\xi}, \delta^{+}_{t,\xi}, \delta^{-}_{t,\xi}, p^{\mathrm{DA}}_{h(t),\xi}, p^{\mathrm{Ru}}_{h(t),\xi}, p^{\mathrm{Rd}}_{h(t),\xi}, p^{\mathrm{RT}}_{q(t),\xi}, e^{\mathrm{s}}_{t,\xi} \in \mathbb{R}^{+}_{0} \quad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{12}$$

Objective function

The objective function in the first stage of scenario-based optimization models is:

$$Z_{\xi} = \sum_{h=h^{\mathrm{DA}}+1}^{h\left(\left|\mathcal{T}^{\mathrm{DA}}\right|\right)} \left(\lambda^{\mathrm{DA}}_{h,\xi} p^{\mathrm{DA}}_{h,\xi} + \lambda^{\mathrm{Rup}}_{h,\xi} p^{\mathrm{Rup}}_{h,\xi} + \lambda^{\mathrm{Rdn}}_{h,\xi} p^{\mathrm{Rdn}}_{h,\xi}\right) \Delta^{60} +$$

$$\sum_{q=q^{\mathrm{RT}}+1}^{q\left(\left|\mathcal{T}^{\mathrm{DA}}\right|\right)} \lambda^{\mathrm{RT}}_{q,\xi} p^{\mathrm{RT}}_{q,\xi} \Delta^{15} - \sum_{t=1}^{\left|\mathcal{T}^{\mathrm{DA}}\right|} \left(\lambda^{\mathrm{imb}} \left|p^{\mathrm{r}}_{t,\xi} - p^{\mathrm{pv}}_{t,\xi} - p^{\mathrm{d}}_{t,\xi} + p^{\mathrm{c}}_{t,\xi}\right|\right) \Delta^{1} \quad \forall \xi \in \Xi \tag{13}$$

please note that the bids to be submitted in the DA market are unique for all scenarios, this was not included in formulations to simplify notation. Nevertheless, this is mathematically equivalent to state:

$$p^{\mathrm{DA}}_{h,\xi} = p^{\mathrm{DA}}_{h,\xi+1} \qquad \forall h \in \{h^{\mathrm{DA}} + 1, ..., h^{\mathrm{DA}} + 24\},\ \forall \xi \in \{1, ..., |\Xi| - 1\} \tag{14}$$

$$p^{\mathrm{Rup}}_{h,\xi} = p^{\mathrm{Rup}}_{h,\xi+1} \qquad \forall h \in \{h^{\mathrm{DA}} + 1, ..., h^{\mathrm{DA}} + 24\},\ \forall \xi \in \{1, ..., |\Xi| - 1\} \tag{15}$$

$$p^{\mathrm{Rdn}}_{h,\xi} = p^{\mathrm{Rdn}}_{h,\xi+1} \qquad \forall h \in \{h^{\mathrm{DA}} + 1, ..., h^{\mathrm{DA}} + 24\},\ \forall \xi \in \{1, ..., |\Xi| - 1\} \tag{16}$$

Initial conditions

The initial conditions included in the first stage of scenario-based optimization models are:

$$p^{\mathrm{DA}}_{h,\xi} = \overline{p}^{\mathrm{DA}}_{h} \qquad \forall h \in \{1, ..., h^{\mathrm{DA}}\},\ \forall \xi \in \Xi \tag{17}$$
$$p^{\mathrm{Rup}}_{h,\xi} = \overline{p}^{\mathrm{Rup}}_{h} \qquad \forall h \in \{1, ..., h^{\mathrm{DA}}\},\ \forall \xi \in \Xi \tag{18}$$
$$p^{\mathrm{Rdn}}_{h,\xi} = \overline{p}^{\mathrm{Rdn}}_{h} \qquad \forall h \in \{1, ..., h^{\mathrm{DA}}\},\ \forall \xi \in \Xi \tag{19}$$
$$p^{\mathrm{RT}}_{q,\xi} = \overline{p}^{\mathrm{RT}}_{q} \qquad \forall q \in \{1, ..., q^{\mathrm{RT}}\},\ \forall \xi \in \Xi \tag{20}$$

where $\overline{p}^{\mathrm{DA}}_{h}$, $\overline{p}^{\mathrm{Rup}}_{h}$, $\overline{p}^{\mathrm{Rdn}}_{h}$, and $\overline{p}^{\mathrm{RT}}_{q}$ are previously made bidding decisions at current time (Π^{m}_{t}).

Constraints

The constraints included in the first stage of scenario-based optimization models are:

$$p^{\mathrm{g}}_{t,\xi} = p^{\mathrm{pv}}_{t,\xi} + p^{\mathrm{d}}_{t,\xi} - p^{\mathrm{c}}_{t,\xi} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{21}$$
$$p^{\mathrm{r}}_{t,\xi} = p^{\mathrm{g}}_{t,\xi} + \delta^{+}_{t,\xi} - \delta^{-}_{t,\xi} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{22}$$
$$p^{\mathrm{r}}_{t,\xi} = p^{\mathrm{DA}}_{h(t),\xi} + b^{+}_{t,\xi} p^{\mathrm{Ru}}_{h(t),\xi} - b^{-}_{t,\xi} p^{\mathrm{Rd}}_{h(t),\xi} + p^{\mathrm{RT}}_{q(t),\xi} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{23}$$
$$e^{\mathrm{s}}_{t,\xi} - e^{\mathrm{s}}_{ini} = \left(\eta^{\mathrm{c}} p^{\mathrm{c}}_{t,\xi} - \frac{p^{\mathrm{d}}_{t,\xi}}{\eta^{\mathrm{d}}}\right)\Delta^{1} \qquad \forall t \in \{1\},\ \forall \xi \in \Xi \tag{24}$$
$$e^{\mathrm{s}}_{t+1,\xi} - e^{\mathrm{s}}_{t,\xi} = \left(\eta^{\mathrm{c}} p^{\mathrm{c}}_{t+1,\xi} - \frac{p^{\mathrm{d}}_{t+1,\xi}}{\eta^{\mathrm{d}}}\right)\Delta^{1} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{25}$$
$$\check{e}^{\mathrm{s}} \leq e^{\mathrm{s}}_{t,\xi} \leq \hat{e}^{\mathrm{s}} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{26}$$
$$0 \leq p^{\mathrm{c}}_{t,\xi} \leq \hat{p}^{\mathrm{s}}\ ,\ 0 \leq p^{\mathrm{d}}_{t,\xi} \leq \hat{p}^{\mathrm{s}} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{27}$$
$$\check{\alpha} \leq p^{\mathrm{DA}}_{h(t),\xi} \leq \hat{\alpha} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{28}$$
$$\check{\alpha} \leq p^{\mathrm{DA}}_{h(t),\xi} + p^{\mathrm{RT}}_{h(t),\xi} \leq \hat{\alpha} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{29}$$
$$0 \leq p^{\mathrm{Ru}}_{h(t),\xi} \leq \hat{\beta} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{30}$$
$$0 \leq p^{\mathrm{Rd}}_{h(t),\xi} \leq \hat{\gamma} \qquad \forall t \in \mathcal{T}^{\mathrm{DA}},\ \forall \xi \in \Xi \tag{31}$$

where e^{s}_{ini} is the ESS' amount of stored energy at current time.

Second stage

Parameters and sets

The parameters needed to formulate the second stage of scenario-based optimization models are:

$$h^{\mathrm{DA}} = 10\ ,\ q^{\mathrm{RT}} = 4\ ,\ \mathcal{T}^{\mathrm{DA}} = \left\{1, ..., \frac{h^{\mathrm{RT}}}{\Delta^{1}}\right\}\ ,\ \Xi = \{1, ..., 10\} \tag{32}$$

where h^{DA} and q^{RT} indicates the time steps where bidding decisions where already made for the DA and RT market, respectively. $\mathcal{T}^{\mathrm{RT}}$ and Ξ are the sets containing optimization's time horizon in 1-min resolution and generated scenarios indexes, respectively.

Decision variables

The decision variables considered in the second stage of scenario-based optimization models are:

$$p^{\mathrm{g}}_{t,\xi}, p^{\mathrm{d}}_{t,\xi}, p^{\mathrm{c}}_{t,\xi}, p^{\mathrm{r}}_{t,\xi}, \delta^{+}_{t,\xi}, \delta^{-}_{t,\xi}, p^{\mathrm{DA}}_{h(t),\xi}, p^{\mathrm{Ru}}_{h(t),\xi}, p^{\mathrm{Rd}}_{h(t),\xi}, p^{\mathrm{RT}}_{q(t),\xi}, e^{\mathrm{s}}_{t,\xi} \in \mathbb{R}^{+}_{0} \quad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \tag{33}$$

$$z^{\mathrm{n}_1}_{t,\xi}, z^{\mathrm{n}_2}_{t,\xi}, z^{\mathrm{p}_1}_{t,\xi}, z^{\mathrm{p}_2}_{t,\xi} \in \{0,1\} \quad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \tag{34}$$

Objective function

The objective function in the second stage scenario-based optimization models is:

$$Z_{\xi} = \sum_{h=h^{\mathrm{DA}}+1}^{h\left(\left|\mathcal{T}^{\mathrm{RT}}\right|\right)} \left(\lambda^{\mathrm{DA}}_{h,\xi} p^{\mathrm{DA}}_{h,\xi} + \lambda^{\mathrm{Rup}}_{h,\xi} p^{\mathrm{Rup}}_{h,\xi} + \lambda^{\mathrm{Rdn}}_{h,\xi} p^{\mathrm{Rdn}}_{h,\xi}\right) \Delta^{60} +$$

$$\sum_{q=q^{\mathrm{RT}}+1}^{q\left(\left|\mathcal{T}^{\mathrm{RT}}\right|\right)} \lambda^{\mathrm{RT}}_{q,\xi} p^{\mathrm{RT}}_{q,\xi} \Delta^{15} - \sum_{t=1}^{\left|\mathcal{T}^{\mathrm{RT}}\right|} \left(\lambda^{\mathrm{imb}} \left|p^{\mathrm{r}}_{t,\xi} - p^{\mathrm{pv}}_{t,\xi} - p^{\mathrm{d}}_{t,\xi} + p^{\mathrm{c}}_{t,\xi}\right|\right) \Delta^{1} \quad \forall \xi \in \Xi \tag{35}$$

please note that the bids to be submitted in the RT market are unique for all scenarios, this was not included in formulations to simplify notation. Nevertheless, this is mathematically equivalent to state:

$$p^{\mathrm{RT}}_{q,\xi} = p^{\mathrm{RT}}_{q,\xi+1} \qquad \forall q \in \{q^{\mathrm{RT}}+1, ..., q^{\mathrm{RT}}+4\},\ \forall \xi \in \{1, ..., |\Xi|-1\} \tag{36}$$

Initial conditions

The initial conditions included in the second stage of scenario-based optimization models are:

$$p^{\mathrm{DA}}_{h,\xi} = \overline{p}^{\mathrm{DA}}_{h} \qquad \forall h \in \{1, ..., h^{\mathrm{DA}}\},\ \forall \xi \in \Xi \tag{37}$$

$$p^{\mathrm{Rup}}_{h,\xi} = \overline{p}^{\mathrm{Rup}}_{h} \qquad \forall h \in \{1, ..., h^{\mathrm{DA}}\},\ \forall \xi \in \Xi \tag{38}$$

$$p^{\mathrm{Rdn}}_{h,\xi} = \overline{p}^{\mathrm{Rdn}}_{h} \qquad \forall h \in \{1, ..., h^{\mathrm{DA}}\},\ \forall \xi \in \Xi \tag{39}$$

$$p^{\mathrm{RT}}_{q,\xi} = \overline{p}^{\mathrm{RT}}_{q} \qquad \forall q \in \{1, ..., q^{\mathrm{RT}}\},\ \forall \xi \in \Xi \tag{40}$$

where $\overline{p}^{\mathrm{DA}}_{h}$, $\overline{p}^{\mathrm{Rup}}_{h}$, $\overline{p}^{\mathrm{Rdn}}_{h}$, and $\overline{p}^{\mathrm{RT}}_{q}$ are previously made bidding decisions at current time (Π^{m}_{t}).

Constraints

The constraints included in the second stage of scenario-based optimization models are:

$$p^{\mathrm{g}}_{t,\xi} = p^{\mathrm{pv}}_{t,\xi} + p^{\mathrm{d}}_{t,\xi} - p^{\mathrm{c}}_{t,\xi} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (41)$$

$$p^{\mathrm{r}}_{t,\xi} = p^{\mathrm{g}}_{t,\xi} + \delta^{+}_{t,\xi} - \delta^{-}_{t,\xi} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (42)$$

$$p^{\mathrm{r}}_{t,\xi} = p^{\mathrm{DA}}_{h(t),\xi} + b^{+}_{t,\xi} p^{\mathrm{Ru}}_{h(t),\xi} - b^{-}_{t,\xi} p^{\mathrm{Rd}}_{h(t),\xi} + p^{\mathrm{RT}}_{q(t),\xi} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (43)$$

$$e^{\mathrm{s}}_{t,\xi} - e^{\mathrm{s}}_{ini} = \Big(\eta^{\mathrm{c}} p^{\mathrm{c}}_{t,\xi} - \frac{p^{\mathrm{d}}_{t,\xi}}{\eta^{\mathrm{d}}}\Big)\Delta^{1} \qquad \forall t \in \{1\},\ \forall \xi \in \Xi \qquad (44)$$

$$e^{\mathrm{s}}_{t+1,\xi} - e^{\mathrm{s}}_{t,\xi} = \Big(\eta^{\mathrm{c}} p^{\mathrm{c}}_{t+1,\xi} - \frac{p^{\mathrm{d}}_{t+1,\xi}}{\eta^{\mathrm{d}}}\Big)\Delta^{1} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (45)$$

$$z^{\mathrm{n}_1}_{t,\xi} + z^{n_2}_{t,\xi} \leq 1 \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (46)$$

$$z^{\mathrm{n}_1}_{t,\xi} \leq \frac{p^{\mathrm{c}}_{t,\xi}}{\hat{p}^{\mathrm{s}}} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (47)$$

$$z^{\mathrm{n}_2}_{t,\xi} \leq \frac{e^{\mathrm{s}}_{t,\xi}}{\hat{e}^{\mathrm{s}}} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (48)$$

$$\delta^{-}_{t,\xi} \leq M \cdot (z^{\mathrm{n}_1}_{t,\xi} + z^{\mathrm{n}_2}_{t,\xi}) \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (49)$$

$$z^{\mathrm{p}_1}_{t,\xi} + z^{p_2}_{t,\xi} \leq 1 \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (50)$$

$$z^{\mathrm{p}_1}_{t,\xi} \leq \frac{p^{\mathrm{d}}_{t,\xi}}{\hat{p}^{\mathrm{s}}} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (51)$$

$$z^{\mathrm{p}_2}_{t,\xi} \leq \frac{\hat{e}^{\mathrm{s}} - e^{\mathrm{s}}_{t,\xi}}{\hat{e}^{\mathrm{s}} - \check{e}^{\mathrm{s}}} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (52)$$

$$\delta^{+}_{t,\xi} \leq M \cdot (z^{\mathrm{p}_1}_{t,\xi} + z^{\mathrm{p}_2}_{t,\xi}) \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (53)$$

$$\check{e}^{\mathrm{s}} \leq e^{\mathrm{s}}_{t,\xi} \leq \hat{e}^{\mathrm{s}} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (54)$$

$$0 \leq p^{\mathrm{c}}_{t,\xi} \leq \hat{p}^{\mathrm{s}}\ ,\ 0 \leq p^{\mathrm{d}}_{t,\xi} \leq \hat{p}^{\mathrm{s}} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (55)$$

$$\check{\alpha} \leq p^{\mathrm{DA}}_{h(t),\xi} \leq \hat{\alpha} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (56)$$

$$\check{\alpha} \leq p^{\mathrm{DA}}_{h(t),\xi} + p^{\mathrm{RT}}_{h(t),\xi} \leq \hat{\alpha} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (57)$$

$$0 \leq p^{\mathrm{Ru}}_{h(t),\xi} \leq \hat{\beta} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (58)$$

$$0 \leq p^{\mathrm{Rd}}_{h(t),\xi} \leq \hat{\gamma} \qquad \forall t \in \mathcal{T}^{\mathrm{RT}},\ \forall \xi \in \Xi \qquad (59)$$

where $z^{\mathrm{p}_1}_{t,\xi}$, $z^{\mathrm{p}_2}_{t,\xi}$, $z^{\mathrm{n}_1}_{t,\xi}$, and $z^{\mathrm{n}_2}_{t,\xi}$ are ancillary binary variables to ensure (4) and M is a large enough number. In order to avoid high computational efforts, (46) to (53) are relaxed after 3 h since second stage optimization starting time, i.e., $t \geq 3/\Delta^{1}$.