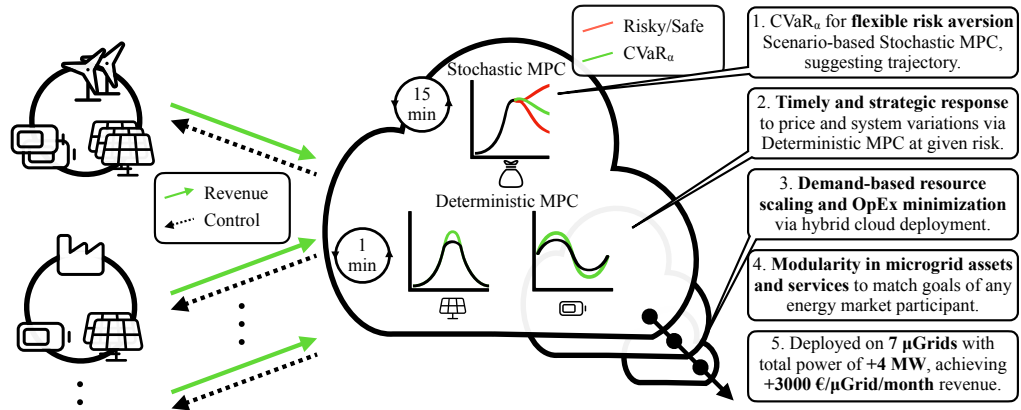


Graphical Abstract

Predictive risk-aware control for microgrids: Operation of a revenue-generating energy management system

Tereza Ábelová, Marek Wadinger, Michal Kvasnica



Highlights

Predictive risk-aware control for microgrids: Operation of a revenue-generating energy management system

Tereza Ábelová, Marek Wadinger, Michal Kvasnica

- Scenario-based predictive control with adjustable risk aversion for microgrids.
- Two-stage procedure combining stochastic and deterministic MPC for efficiency.
- Scalable, modular software architecture for commercial microgrid deployments.
- Validated in operation, achieving over 3000 €/month through revenue stacking.

Predictive risk-aware control for microgrids: Operation of a revenue-generating energy management system

Tereza Ábelová*, Marek Wadinger, Michal Kvasnica

Department of Information Engineering and Process Control, Slovak University of Technology in Bratislava, Radlinského 9, 81237, Bratislava, Slovakia

Abstract

Economically sustainable operation of microgrid systems helps with transition of energy sector towards renewable energy sources. This paper presents a comprehensive energy management framework for microgrids, integrating advanced control methodologies with practical deployment considerations. The proposed framework employs a layered control structure, with a focus on the energy management layer that utilizes a scenario-based model predictive control (MPC) approach to address uncertainties using a risk measure in microgrid systems. A novel two-stage MPC method is introduced to enhance computational efficiency, combining stochastic and deterministic MPC. The software architecture is designed for scalability, consciously manages cloud-based resources for computationally demanding tasks while maintaining local control for real-time operations. The framework supports modular microgrid composition, revenue stacking, and delivers profit generation by optimizing multiple revenue streams, including energy arbitrage, peak-shaving, and imbalance settlement. The framework is validated through a case study of a commercial microgrid with photovoltaic power plant accompanied by a large-scale battery energy storage, demonstrating significant economic benefits and operational reliability. The results highlight the framework's competence to make microgrid operation profitable.

Keywords: Microgrid, Battery energy storage, Energy management system, Model predictive control, Software architecture, Profit optimization

*Corresponding author

Email address: `tereza.abelova@stuba.sk` (Tereza Ábelová)

1. Introduction

The energy management systems (EMS) are central to the operation of microgrids, serving as the interface between process control and information engineering. This paper focuses on two key aspects of EMS design: (1) the development of control algorithms to optimize energy flows and meet operational objectives, and (2) the practical implementation of the EMS software architecture to ensure reliable and scalable deployment. Model predictive control (MPC) is the most suitable approach for addressing challenges like cost optimization, renewable energy integration, and uncertainty management while adhering to system constraints. At the same time, EMS deployment requires addressing practical aspects regarding execution platform, data handling, or real-time communication. By combining these perspectives, the proposed framework connects theoretical advances in control with the practical requirements of operational microgrids, ensuring adaptability to dynamic environments and compatibility with market and grid needs.

Microgrids that integrate battery energy storage systems (BESS) and renewable energy sources (RES) are vital for the transition of the energy sector. As renewables expand and increase variability in power networks, effective microgrid control is essential for maintaining balance and supporting broader grid stability. Economic aspects are also crucial, ensuring that microgrid operation is not only sustainable but financially viable for operators and stakeholders. An important aspect of EMS offered as Microgrid Control as a Service (MaaS) is to be a profit-generating service provider. Managing multiple unique microgrids requires modularity and scalability, with benefits shared between both the microgrid and the provider. In addition to operational costs, the control architecture and computational resources must be considered so that the EMS remains profitable while keeping computing costs in check.

1.1. *Current state of research*

The integration of renewable energy sources and battery storage systems into microgrids has been extensively studied to enhance sustainability and operational efficiency. Model predictive control has emerged as a prominent strategy for managing energy flows within these systems, particularly under the constrained environment, multiple control objectives, and uncertainties inherent in RES outputs and load demands.

Several studies have explored MPC-based approaches for microgrid energy management. Deterministic MPC methods, which do not explicitly consider uncertainty, have been widely studied. For instance, in [1], an economic MPC framework is proposed for optimizing energy exchange between a residential microgrid and the main grid. The objective function consists of minimizing electricity costs and battery usage costs. However, disturbances in generation and load are only handled within the receding horizon framework without explicitly modelling uncertainty. Moreover, the approach is evaluated only in simulations with a long sampling period (1 hour), making it impractical for real-time operation in microgrids with power fluctuations occurring on a second-to-minute scale. Similar deterministic approach was taken in [2], where problem combines two objectives: price arbitrage and schedule deviations, while incorporating battery degradation model in energy management.

Robust optimization techniques aim to handle uncertainties conservatively. In [3], a multi-stage robust optimization approach is proposed for shared energy storage management across multiple microgrids. The method ensures operational reliability while balancing cost efficiency. However, robustness can lead to overly conservative decisions, potentially reducing economic benefits. Unlike conventional robust optimization, risk-averse formulations such as [4] employ Conditional Value at Risk (CVaR) to provide an adjustable level of conservatism. This approach minimizes exposure to extreme financial losses rather than purely optimizing for the worst case, allowing for a more balanced trade-off between risk and profitability. Nonetheless, its implementation updates control decisions only every 30 minutes.

Stochastic MPC methods explicitly incorporate forecast uncertainty into decision-making through scenario-based optimization. In [5], a stochastic MPC approach is proposed for grid-connected microgrids with electric vehicles (EVs) with uncertainty in number of EVs and their state of charge. The study approximates the stochastic features by using a multi-scenario sampling approach. However, like many existing frameworks, this study employs a 15-minute control update, which is insufficient for microgrids providing energy services such as imbalance settlement. Instead of uncertainty in power generation/consumption, work in [5] considers uncertainty in battery state and timing of electric vehicles (EV) within microgrid and proposes a two-layer multi-scenario MPC, where the upper layer optimizes the overall economics and cost reduction, the lower layer focuses on power allocation for individual EVs. Computational and economic efficiency is demonstrated only in simu-

lations. A two-layered control hierarchy is also adopted in [6], with a slower high-level layer computing optimal reference for a faster low-level layer addressing uncertainty in load and PV power. The proposed scenario-based MPC verified in a lab-scale nanogrid outperforms both a chance-constrained MPC and a rule-based EMS.

Several studies integrate multi-objective and multi-layer control frameworks. In [7], a multi-objective stochastic optimization model is presented for microgrids with energy storage and diesel generators, transforming the problem into a single-objective optimization using a weighted sum approach, but also working with 1-hour average powers. A two-layer stochastic multi agent framework is proposed in [8], where a high-level control unit manages nominal forecasts of photovoltaic production and load consumption while a low-level stochastic MPC operating with higher frequency refines optimal decisions based on real-time disturbances. Similarly, in [9], stochastic MPC with CVaR is employed to optimize battery storage participation in electricity markets while balancing profitability, financial risks, and degradation costs.

A more market-oriented perspective is considered in [10, 11], where a risk-averse aggregator utilizes CVaR-based optimization formulated as a mixed-integer linear program (MILP) to provide multiple services, including reserve markets and imbalance settlements. However, the study focuses on storage-only systems, without considering renewables or uncertainty in power generation. In [12], a standalone storage is optimized for price arbitrage under uncertainty using both robust and chance-constrained decision-making. While this approach enhances market adaptability, it does not integrate microgrid operation as a whole, limiting its applicability in practical settings.

The objective of the microgrid control problem is most commonly formulated as a cost reduction, which includes minimizing energy procurement costs from the grid complemented with battery degradation expenses [12, 3, 1]. Other key objectives include reducing peak demands and flattening the load profile [13, 14], minimizing the grid bidirectional exchange regardless of the price [15], or minimizing imbalance deviations. However, when addressing the imbalance, many studies overlook the fact that not all energy deviations incur penalties. Few approaches actively take advantage of the imbalance pricing by strategically positioning the system out of balance, as seen in [16, 9, 10]. In addition to economic considerations, some works incorporate environmental impact by integrating pollutant treatment cost models into the optimization framework [7, 17].

Despite these advancements, existing approaches often face limitations in modularity, scalability, and practical implementation under real-world conditions. Results are demonstrated in simulations only or with experiments in a laboratory [6], thus not facing problems with deployment and long-term operation. Many frameworks are designed for a particular application with a single revenue stream and with a pre-defined microgrid composition of selected assets, lacking generality and versatility. The business perspective of such solutions is also an attribute we find important to consider. Additionally, the computational complexity of some stochastic and robust MPC methods can impede their real-time execution, especially in systems with limited processing capabilities.

Extensive exploration has been conducted on deep learning techniques, where the objective is not to surpass MPC but to imitate its behavior through learning to overcome the computational burden, demonstrated with a single revenue stream in [15, 18]. Recent work employs deep learning for forecasting and a MILP model for real multiple revenue stream cases in [19], though the hourly control updates limit applicability for current 15 minute intra-day trading and microgrids requiring real-time control to meet the strategic objectives.

Energy management systems provide the framework for deploying advanced control strategies like MPC in real-time microgrid control. While control methodologies, economic optimization, and cloud-based infrastructure have been widely studied independently, their integration presents significant challenges. The practical aspect of large-scale EMS deployment, creating cost-efficient architectures that provide long-term value across various microgrids, remains an open challenge.

Conventional local deployment approaches require dedicated control units for each microgrid, enabling real-time control with minimal latency. Nevertheless, this approach introduces tradeoff between advanced control capabilities and significant upfront hardware investments [20]. While local deployment ensures operational autonomy in islanded mode, the economic burden of distributed hardware limits its viability for large-scale microgrid networks, which are subject to energy market fluctuations.

Hierarchical architectures are widely adopted due to their ability to separate long-term planning from real-time operational control [21]. The proposed EMS comprises a two-level hierarchical control structure, with local controllers at microgrid level, and a supervisory control at the main-grid level [22]. A two-layer MPC approach has been proposed for hydroelectric

power management [23], demonstrating how predictive control strategies can be implemented in real-time microgrid control. Similarly, Wu et al. [24] present a hierarchical online EMS using fuzzy logic-based controllers for residential microgrids with hybrid hydrogen-electricity storage systems, where decision-making is decoupled by different operating frequencies to handle both seasonal and daily energy allocation.

Hybrid architectures combine predictive optimization with reliability. Rule-based logic with reinforcement learning is used in multi-stack and multi-energy configurations [25]. Neural network-based EMS framework have also emerged for real-time scheduling and control across multiple timescales, effectively managing operational complexity in integrated systems [26]. Recent co-optimization frameworks demonstrate how sizing, scheduling, and EMS logic can be aligned across physical and virtual layers, improving maintainability and economic performance in heterogeneous deployments [27]. Despite these advancements, existing EMS frameworks rarely address deployment overheads, cost-efficiency, or the architectural scalability required for revenue-generating microgrid networks.

While centralized architectures offer superior optimization capabilities, distributed and decentralized approaches have also been proposed to enhance scalability. A decentralized market-based approach is presented in [28], where real-time optimization dynamically adjusts microgrid operations based on evolving forecasts. However, this approach assumes that frequent forecast updates are sufficient to address uncertainty, overlooking scenarios where poor forecast quality could lead to significant economic losses regardless of update frequency. An energy management combining day-ahead planning for individual microgrids and cooperative optimization across the microgrid cluster is proposed in [29].

The shift toward cloud-based architectures has further influenced EMS implementations. Cloud computing enables scalable optimization and asset flexibility while minimizing capital expenditures (CapEx) for microgrid operators [30]. The synchronization and horizontal scalability to the number of assets and operators is achieved by clustering brokers and specifying communication streams reflecting the desired level of isolation and redundancy [31]. Furthermore, modular EMS architectures enhance adaptability across different microgrid configurations. The work in [32] demonstrates the benefits of a modular software framework, allowing efficient replication across different building energy systems, while also ensuring robustness against modeling errors.

1.2. Contributions

Despite extensive prior work, a clear gap remains: there is no end-to-end, deployable EMS framework that simultaneously (i) integrates risk-aware, revenue-stacking MPC suitable for heterogeneous microgrids, (ii) guarantees sub-minute receding-horizon updates using non-proprietary solvers, (iii) scales across diverse asset compositions via modular abstractions, and (iv) explicitly optimizes operational expenditures through dynamic cloud resource allocation with quantified cost models. Existing studies typically cover only subsets of these needs—e.g., control algorithms without deployment economics, scalable cloud stacks without real-time MPC guarantees, or simulation-only validations without multi-site operational evidence.

A particularly important shortcoming of prior MPC-based EMS approaches is that they almost exclusively generate day-ahead schedules with coarse temporal granularity. These schedules assume constant power delivery over long intervals, which ignores the fact that generation and demand fluctuate almost instantaneously. As a result, existing methods may appear optimal in aggregated energy terms but lack validation how it would perform in dynamic operating environment.

To close this gap, we propose a hybrid EMS architecture that couples a two-stage, CVaR-based MPC with a cloud-edge deployment that dynamically allocates compute to workload intensity (FaaS for frequent low-latency dMPC, CaaS for infrequent higher-duration sMPC). This design achieves sub-minute control with computational feasibility at commercial scale while minimizing OpEx. The framework has been deployed across several industrial microgrids, demonstrating reliable real-time operation, scalable multi-site provisioning, and measurable profitability for operators and providers.

The key contributions are namely as follows:

1. **Profit-generating control scheme for microgrid energy management** incorporates the model of the broader energy ecosystem, encompassing power networks, energy markets, balance markets, and energy suppliers, to identify and exploit economic opportunities with modular revenue stacking. Uniquely, our control strategy actively generates profit by exploiting multiple revenue streams including imbalance settlement opportunities that are often overlooked in existing literature.
2. **Modular framework for versatile microgrid compositions:** extends to various functionalities such as price arbitrage, peak-shaving,

and imbalance settlement. Unlike existing work, our framework supports easy configuration of functionalities and assets, which can be added or removed without altering the overall structure, allowing the EMS to be adapted to different setups ensuring flexibility and scalability.

3. **Novel risk-aware MPC methodology for uncertainty management:** addresses inherent uncertainties in renewable generation, load demands, and market conditions, while enhancing the computational efficiency compared to robust MPC. Real-time decision-making remains feasible within the complex microgrid environments while maintaining minute-level response.
4. **Software architecture with cost consideration:** resolves financial challenges of implementing EMS as a service in real-world microgrids. We address communication between control layers and real-time data handling, while considering the scalability and operational costs of EMS provisioning across multiple microgrids, ensuring its economic viability as a MaaS.
5. **Deployment and validation in commercial and industrial facilities:** demonstrate effectiveness of our solution in managing multiple mid- to large-scale energy assets, with results showing significant economic benefits arising from the developed energy management system, proving practical viability of this contribution.

2. Microgrid energy management system

2.1. Outline

In this paper, we define a microgrid as a behind-the-meter energy system—a local, small-scale network of electricity sources and loads that maintains connection to the main power grid. Microgrids can incorporate various power generation sources, including renewable energy systems (such as photovoltaic panels and wind turbines), conventional generators, energy storage systems, and both controllable and non-controllable loads. Their primary function is to manage energy production, consumption, and storage in a manner that ensures reliable and efficient operation while minimizing costs.

The operation of microgrids is predominantly driven by economic considerations—microgrids interact with energy and balance markets, energy suppliers and grid operators. This environment enables revenue stacking from various sources. These streams include price arbitrage (using more

energy when it is inexpensive and less when it is costly), peak shaving (reducing peak loads to lower demand charges), renewable energy utilization and self-consumption (maximizing on-site renewable energy use), and imbalance settlement (minimizing deviations from schedules or capitalizing on imbalance pricing).

Effective management of energy assets within the microgrid is the responsibility of the operator, who must optimize operations to fulfil all required functionalities. This involves decision-making across multiple timescales. For instance, participation in energy markets requires planning several days in advance, whereas the rapid dynamics of power electronics require real-time control. A similar layered approach to decision-making aligns with the methodologies discussed in [33].

2.2. Control hierarchy

The comprehensive energy management system is structured into four decision-making layers, all summarized in Table 1. Each layer is responsible for specific tasks, operates at different timescales, and considers different scopes. Output from one layer serves as the input to the next. The control hierarchy with data flows is illustrated in Figure 1.

2.2.1. Business layer

The uppermost layer, the business layer, operates on a timescale of months to years. This layer determines the goals (e.g., saving costs or reducing carbon emissions) and functionalities (e.g., price arbitrage or maximizing self-consumption) the microgrid should pursue, considering external conditions, contractual agreements with energy providers, and the composition of the microgrid. Business layer can make strategic decisions about the microgrid structure (e.g. BESS nominal capacity and maximum power output, E_{nom} and \bar{P}_{bat}). Decisions made at this level are subject to approval by a business entity such as a customer, operator, or stakeholder. The outputs of this layer define the objectives for the next—the energy management layer.

2.2.2. Energy management layer

The energy management layer focuses on real-time decisions regarding the energy quantities allocated to each group of assets. This layer operates with a predictive horizon of hours to days, considering energy quantities with intermediate granularity. The model predictive control is preferable strategy due to its ability to optimize a schedule over time while respecting operational

Table 1: Control hierarchy in the microgrid energy management system, output from one layer serves as the input to the next. Main scope of this paper is the energy management layer.

	Business layer	Energy management layer	Power management layer	Device control layer
Role	Define goals and functionalities	Optimize energy flows	Distribute power setpoints, SoC balancing, fallback mode	Execute power commands
Inputs	Microgrid parameters, contractual agreements	Forecasts, objectives, real-time measurements	Energy schedule, real-time measurements	Power setpoints from upper layers
Outputs	Objectives for energy management	Optimal energy schedule for each asset group	Power setpoints for individual devices	Voltage and current regulation
Controller	Informed expert decision	Model predictive controller	Rule-based controller	PID controller
Timescale	Days to years	Minutes to hours	Seconds	Milliseconds

constraints to ensure safe operations. MPC can directly integrate models and parameters of various microgrid assets, making it a robust choice for this decision layer. The proposed MPC-based methodology, detailed further in Section 2.4, serves as the foundation for this layer. It aims to allocate energy quantities with a granularity of minutes, typically aligning with 15-minute billing settlement periods. Decisions are computed in real-time with frequent updates, ranging from seconds to a few minutes, depending on computational resources.

This layer processes inputs such as the business layer’s objectives, parameters of microgrid components, current measured states, and forecasted states over the prediction horizon. Its output is an energy schedule for the upcoming hours, consisting of an array of optimal power setpoints (e.g. power demand for all BESS units, P_{bat}). From this array, the first setpoint is forwarded to the secondary control layer for execution.

2.2.3. Power management layer

The role of secondary control layer, or power management level, is to validate and distribute the setpoints received from the energy management layer. It is necessary for this layer to operate with a faster rate than the energy management layer, typically in the range of seconds. Tasks at this layer are often handled with the rule-based controllers. Operating on a timescale of seconds, this layer addresses fast microgrid dynamics and ensures that operational limits are maintained.

Additionally, this secondary layer handles the disaggregation of power setpoints for asset groups into commands for individual units. For example, it may distribute power among N_B BESS units ($P_{\text{bat}}^i \forall i \in \{1, \dots, N_B\}$) or PV arrays with multiple inverters if present. This approach maintains scalability while keeping the computational complexity at the energy management layer constant. The secondary layer also manages the state of charge (SoC) balance among storage units through appropriate power allocation.

Because the EMS relies on a cloud-based architecture, this layer runs on-site and serves as a fail-safe in case of network outages or loss of connectivity. In such events, the system automatically switches to a rule-based fallback mode that continues to enforce safe operation of the microgrid, albeit without the optimality provided by the higher-level MPC. This mechanism ensures that supply-demand balance and protection limits are maintained even under communication failures. The specific design of the reactive control logic is implementation-dependent and therefore not detailed here.

2.2.4. Device control layer

At the device level, individual power commands are executed by primary controllers. These controllers, embedded in the devices, regulate the output voltage and current of power electronic converters, typically using PID control. Operating with millisecond-level resolution, this layer ensures precise adherence to the power references provided by the upper layers.

2.3. Microgrid model

Microgrid components are categorized into four asset groups, which are connected to single microgrid AC bus, the scheme is illustrated in Figure 2:

- **Energy sources:** These units generate power, in this paper the focus is on renewable energy sources such as photovoltaic systems and wind turbines. Their output power, denoted as P_{RES} , is fully dependent

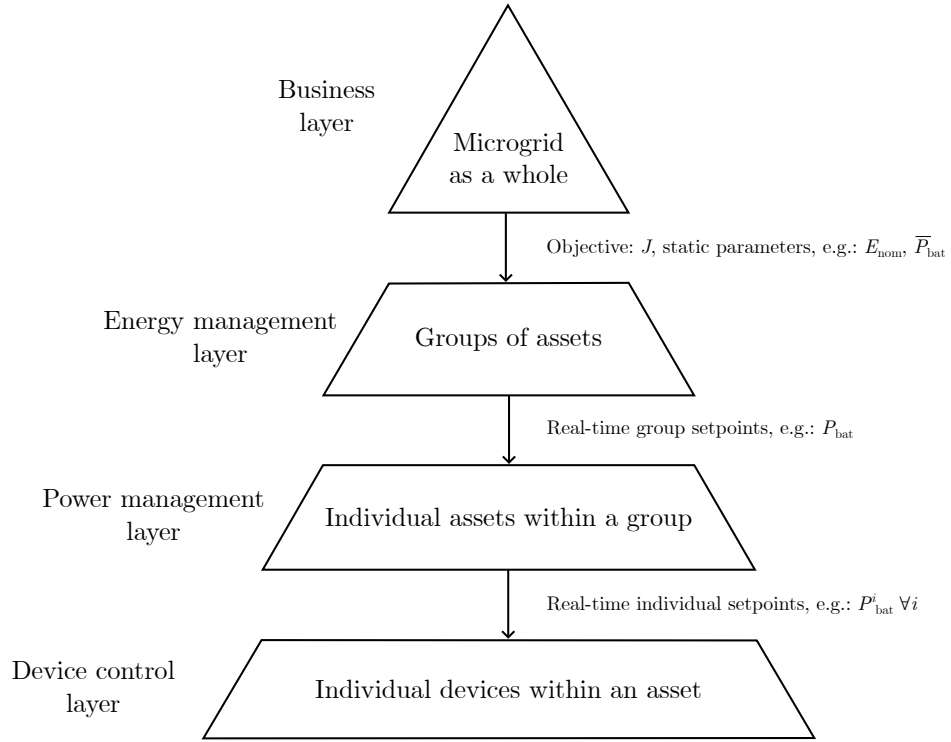


Figure 1: Control hierarchy in the microgrid management system, which operates across four nested scopes.

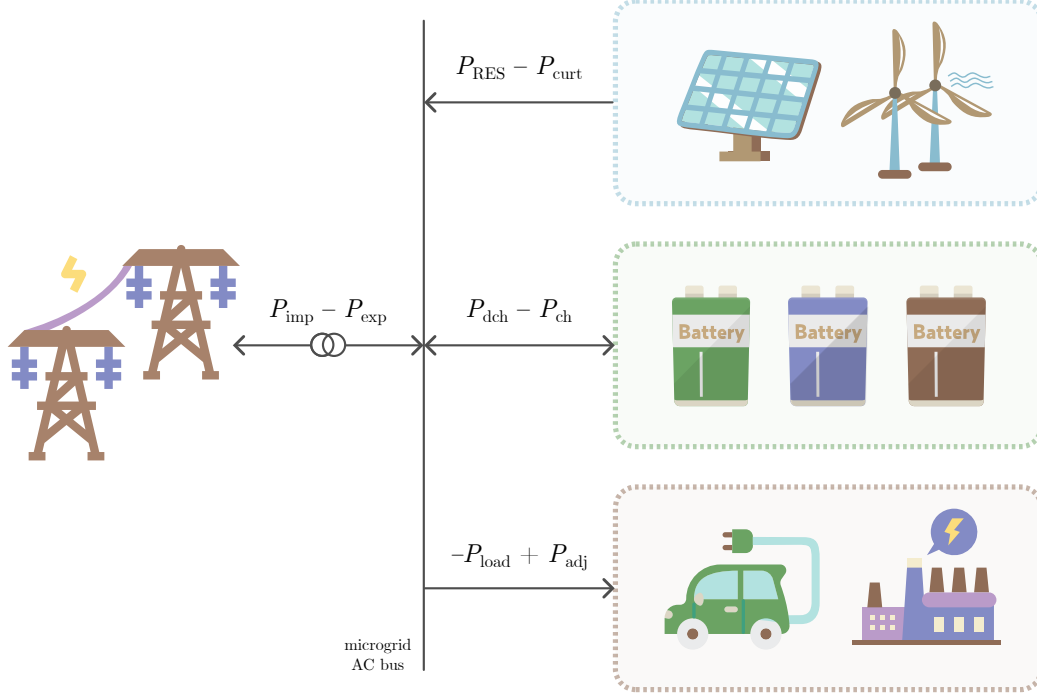


Figure 2: Scheme of a microgrid with four distinct asset groups.

on external conditions. The power injected into the microgrid can be managed, expressed as $P_{\text{RES}} - P_{\text{curt}}$, where P_{curt} represents the curtailing action under our control.

- **Loads:** The energy-consuming units, in microgrids we recognize controllable and non-controllable load, together denoted P_{load} . Controllable load can be adjusted within given limits by the action command P_{adj} . Non-controllable load is assumed to be inflexible and must be fully supplied.
- **Storages:** These units can act as power consumers when charging (P_{ch}) or power suppliers when discharging (P_{dch}). In addition to power variable, they are also described by the internal state of the energy level E .
- **Utility grid:** This is the connection between the microgrid and the main power network. Grid import (P_{imp}) represents power taken by

the microgrid, while grid export (P_{exp}) represents power feed-ins to the main grid.

Each asset group may contain multiple units. For the purposes of the model predictive controller in the energy management layer, these units are aggregated and treated as a single entity.

All components are connected to the microgrid's AC bus. DC-based units, such as PV arrays and battery storages, interface with the AC bus through inverter devices responsible for power conversion. These inverters enable the energy management system to directly manipulate their output power.

2.3.1. Power balance

The microgrid operates under the principle of energy conservation, which requires a balance between production and consumption to be maintained at every time step k . This is expressed by the power balance equation, which binds all power flows within the microgrid and the exchange with the main grid:

$$P_{\text{imp}}(k) + P_{\text{dch}}(k) + P_{\text{RES}}(k) + P_{\text{adj}}(k) = P_{\text{exp}}(k) + P_{\text{ch}}(k) + P_{\text{load}}(k) + P_{\text{curt}}(k). \quad (1)$$

Power generation terms are on the left-hand side, power consumption on the right-hand side. The actual power production from renewable sources is expressed as $P_{\text{RES}}(k) - P_{\text{curt}}(k)$, while the actual load is $P_{\text{load}}(k) - P_{\text{adj}}(k)$. Term P_{adj} is negative when it is intended to increase the consumption. Note that the energy that was intended to be used but was not is interchangeable with the energy produced, so it stands on the left-hand side of the equation.

2.3.2. Energy storage system

Energy storage systems offer the highest flexibility within a microgrid, as they can operate across the full range from energy supply to withdrawal, constrained only by their maximum power capabilities. The state of charge, representing the storage's energy level, imposes additional constraints on the decision variables and dynamically links consecutive time intervals. The discrete battery model describing the evolution of the SoC under charging and discharging operations is expressed as:

$$E(k+1) = E(k) + \Delta T \left(\eta_{\text{ch}} P_{\text{ch}}(k) - \frac{1}{\eta_{\text{dch}}} P_{\text{dch}}(k) \right), \quad (2)$$

where the energy level E is bound:

$$\underline{\beta} E_{\text{avail}} \leq E(k) \leq \overline{\beta} E_{\text{avail}}. \quad (3)$$

The battery model is characterized by the following parameters: maximum charging and discharging power limits $\overline{P}_{\text{ch}}, \overline{P}_{\text{dch}}$, charging and discharging efficiencies $\eta_{\text{ch}}, \eta_{\text{dch}} < 1$, minimum and maximum allowable SoC levels $\underline{\beta}, \overline{\beta} \in [0, 1]$, nominal energy capacity E_{nom} , and state of health $S \in [0, 1]$. State of health (SoH) is a measure of the battery available capacity relative to its nominal value, which degrades over time due to aging and cycling effects. The relation between the nominal, available energy, and SoH is given by $E_{\text{avail}}(k) = S(k) \cdot E_{\text{nom}}$. Because the change in S is a slow process, during the prediction horizon the measure is treated as a constant.

The decision variables P_{ch} and P_{dch} are interdependent, as simultaneous charging and discharging must be avoided. This condition is enforced using a complementary constraint $P_{\text{ch}}(k) \cdot P_{\text{dch}}(k) = 0$, which can be modeled using additional binary variables, $\delta_{\text{ch}}, \delta_{\text{dch}} \in \{0, 1\}$, resulting in a mixed-integer linear program:

$$0 \leq P_{\text{ch}}(k) \leq \delta_{\text{ch}}(k) \overline{P}_{\text{ch}}, \quad (4a)$$

$$0 \leq P_{\text{dch}}(k) \leq \delta_{\text{dch}}(k) \overline{P}_{\text{dch}}, \quad (4b)$$

$$\delta_{\text{ch}}(k) + \delta_{\text{dch}}(k) \leq 1. \quad (4c)$$

In the simple linear programming (LP) formulation, the complementary condition is removed:

$$0 \leq P_{\text{ch}}(k) \leq \overline{P}_{\text{ch}}, \quad (5a)$$

$$0 \leq P_{\text{dch}}(k) \leq \overline{P}_{\text{dch}}. \quad (5b)$$

Our methodology employs both MILP and LP formulations, as discussed in Section 2.4. The often seen assumption that simultaneous charging and discharging will not occur is either not true or valid only under certain conditions [34]. We aim to provide a methodology that utilizes the simplicity of the relaxed formulation while decisive actions are computed using the exact battery model.

2.3.3. Controllable load/source

The microgrid may include assets that offer some, but not full, flexibility in power management by allowing controlled adjustments to their output or

consumption. These assets include renewable energy sources and controllable loads, which can be regulated to achieve operational objectives.

The power output of renewable energy sources P_{RES} represents the maximum available power. Due to operational or grid constraints, a portion of this power can be curtailed, expressed as the reduction variable P_{curt} . Then the actual power in the microgrid is $P_{\text{RES}}(k) - P_{\text{curt}}(k)$. The following constraints ensure the feasibility of P_{curt} :

$$0 \leq P_{\text{curt}}(k) \leq P_{\text{RES}}(k). \quad (6)$$

Here, P_{curt} cannot exceed the power P_{RES} , as it represents a portion of the available renewable generation that can be curtailed.

Similarly, with controllable load, which refers to flexible consumption units whose demand can be reduced or increased when necessary, the variable is P_{load} and the adjustment in controllable load is managed through the variable P_{adj} . Then the realized load demand is the difference $P_{\text{load}}(k) - P_{\text{adj}}(k)$. The constraints for P_{adj} are given by:

$$\underline{P}_{\text{adj}}(k) \leq P_{\text{adj}}(k) \leq \overline{P}_{\text{adj}}(k), \quad (7a)$$

$$P_{\text{adj}}(k) \leq P_{\text{load}}(k), \quad (7b)$$

where limits $\underline{P}_{\text{adj}}, \overline{P}_{\text{adj}}$ keeps the controlled variable aligned with the microgrid capabilities, while second constraint ensures that the reduction does not exceed the load.

2.3.4. Utility grid

Between the microgrid and the main utility grid is bidirectional power flow referred to as import and export. The power exchange is determined as part of the optimization process in the energy management system, balancing the microgrid's internal generation, consumption, and storage to achieve cost efficiency and operational reliability. The imported power P_{imp} is the quantity that the microgrid pays for (unless the prices are negative), while the exported power P_{exp} represents the power supplied to the grid, associated with remuneration (or expense in case of negative prices).

To ensure compliance with grid connection limits and operational constraints, the power exchange with the utility grid is subject to the following bounds:

$$0 \leq P_{\text{imp}}(k) \leq \overline{P}_{\text{imp}} + \epsilon_{\text{imp}}(k), \quad (8a)$$

$$0 \leq P_{\text{exp}}(k) \leq \overline{P}_{\text{exp}} + \epsilon_{\text{exp}}(k), \quad (8b)$$

where \bar{P}_{imp} and \bar{P}_{exp} are the maximum import and export limits, respectively, imposed by the utility grid operator. The limits are softened with non-negative slacks $\epsilon_{\text{imp}}, \epsilon_{\text{exp}}$ to convey that limits are not hard, but the violation is associated with the penalty q_{imp} and q_{exp} for import and export, respectively.

Microgrids which have responsibility for thier imbalance are required to submit their planned energy schedule to the grid operator a day in advance. They are then responsible for any deviations between the scheduled and actual energy withdrawal or injection.

The imbalance settlement—usually considered as an obligation for energy consumers/producers—also offers the opportunities. Imbalance management typically aims to minimize deviations from the scheduled energy profile to avoid penalties, but with a predictive framework it is possible to take the advantage of the imbalance pricing mechanism. The mechanism depends on the interaction between the system imbalance (total imbalance in the balancing area of the utility grid) and the responsible party, here the microgrid. For example, when the utility grid in the balancing area is in surplus, the responsible parties must pay for increasing the surplus but are compensated for being in the shortage. Conversely, when the system is in the shortage, contributions toward surplus are incentivized. The sign and magnitude of the imbalance prices determine the flow of costs or incentives, enabling control strategies to optimize microgrid operations for economic benefit while maintaining the system balance.

The scheduled profile is denoted as P_{DA} , representing the day-ahead diagram, known for every time step k . The deviations from the schedule result in either a power shortage or surplus, defined as follows:

$$P_{\text{short}}(k) = \max(0, P_{\text{imp}}(k) - P_{\text{exp}}(k) - P_{\text{DA}}(k)), \quad (9a)$$

$$P_{\text{long}}(k) = \max(0, P_{\text{DA}}(k) - P_{\text{imp}}(k) + P_{\text{exp}}(k)). \quad (9b)$$

The variable P_{short} quantifies the shortage in energy delivery relative to the schedule, while P_{long} represents the surplus energy delivered beyond the scheduled amount. Both terms are non-negative by definition. Relations defined in (9) are included in the optimization problem in the following linear form:

$$P_{\text{short}}(k) - P_{\text{long}}(k) = P_{\text{imp}}(k) - P_{\text{exp}}(k) - P_{\text{DA}}(k), \quad (10)$$

where only P_{DA} is known, while the other terms are decision variables.

2.3.5. Costs and revenues

The microgrid revenue model is multifaceted, encompassing both the costs and opportunities associated with external power systems, such as balance markets, grid operators, and energy suppliers, as well as the internal costs related to asset operations. The objective is to maximize the total profit defined as a difference between the revenues and costs.

Each action and power flow within the microgrid is associated with a corresponding cost. For power exchange with the utility grid, the costs are c_{imp} , c_{exp} , c_{short} , and c_{long} for energy import, export, shortage, and surplus, respectively. Variable costs over time enable the controller to perform price arbitrage, where low market prices signal that grid has a surplus of energy, motivating increased energy draw at cheaper rates. Peak-shaving functionality and maintaining grid exchange within the limits is achieved by penalizing the violations ϵ_{imp} and ϵ_{exp} .

Storage charging and discharging actions are evaluated with parameters c_{ch} and c_{dch} , accounting for capacity degradation due to cycling. Similarly, actions of curtailing renewable energy and adjusting controllable load are evaluated with costs c_{curt} and c_{adj} . Adjusting cost coefficients allows prioritization of maximizing self-consumption of renewable energy over grid power.

The total profits to be maximized are a linear function of cost coefficients and power flows. This approach allows revenue stacking, with the controller dynamically prioritizing functionalities based on changing conditions.

2.4. Model predictive controller

The widely used model predictive controller serves as the core of the energy management layer for decision-making in this study. In highly volatile environment where forecasts are uncertain, conventional deterministic MPC (dMPC) is not sufficient. A robust MPC, when there is a large uncertainty, is too conservative. Stochastic MPC (sMPC) appears as the best middle course, but it is computationally time-consuming, especially when using non-commercial solvers. These existing approaches fail to achieve the desired tradeoff between risk-awareness and computational speed.

To efficiently address uncertainties while meeting the requirement for fast control, a risk-aware methodology with two-stage optimization process was developed to reduce the computation time. This methodology was built upon our previous work, firstly proposed in [35], where computational times were compared to purely stochastic MPC. In this paper, it has been extended with a more general model that includes controllable load, imbalance settlement

and day-ahead diagram, and updated with parallelization of processes to reduce the computation burden even more.

In general scenario-based stochastic MPC, control actions are determined by solving a single optimization problem that incorporates multiple uncertainty realizations. However, when applied to non-linear models, the computational effort scales poorly with the number of scenarios, making it impractical for real-time applications. To address this scalability challenge, our method employs a two-stage approach: first, a scenario-based stochastic MPC is solved using a relaxed linear model, which significantly reduces computation time while still capturing the essential stochastic properties. From this solution, the scenario corresponding to the desired risk level is identified. In the second stage, a deterministic MPC problem is solved to determine the binding actions using the full non-linear model with the identified single scenario, thereby achieving faster computations while capturing the uncertainty. The second stage is therefore an approximation of stochastic MPC, but computed in shorter time. The scheme in Figure 3 compares the time scales of the discussed approaches, highlighting the efficiency gains of our method.

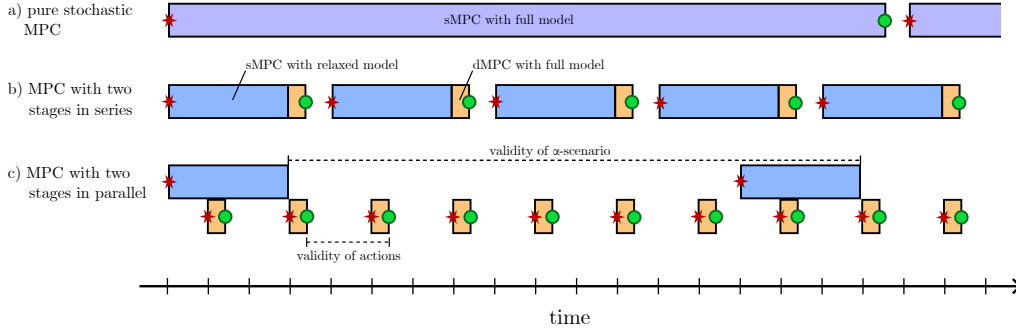


Figure 3: Compared approaches, blocks represent the computation time, red star denotes timestamp of measurement and green circle denotes applying of optimal control actions. This paper employs the approach c), where the time between measurement and applied control, as well as the update frequency is the shortest.

The ultimate problem to be solved is defined as maximization of expected profit in the worst set of scenarios, subjected to the non-linear model of the microgrid. In the next section, both models are derived: the non-linear deterministic MPC and the relaxed stochastic MPC. The full two-stage procedure then combines both approaches to get an approximated solution, but in limited time.

2.4.1. Deterministic MPC formulation

The section starts with defining the variables necessary for MPC notation and builds up towards scenario optimization. Description of the general microgrid instance that allows modularity includes the following arrays:

- **State (E):** the energy level of the storage.
- **Controls (\mathbf{u}):** power flow of controllable assets and power exchange with the main power grid.
- **Uncertainties (\mathbf{w}):** parameters where only the estimation is available—expected power production and consumption.
- **Parameters (\mathbf{p}):** parameters perfectly known, e.g. day-ahead diagram, limits, efficiencies.
- **Costs (\mathbf{c}):** cost coefficients, e.g. grid import costs, battery usage costs.
- **Slacks (\mathbf{e}):** constraint relaxation variables, e.g. for softening the battery state limits, grid limits.
- **Weights (\mathbf{q}):** penalty coefficients for slack variables.

The controller focuses on maximizing the total profit J of microgrid operation over the prediction horizon, split into N steps. The objective function encompasses all operating costs, revenues, and penalties, where each control action (battery usage, power modification of controllable assets, energy flow to/from grid) is associated with its cost and the sign (positive for revenues, negative for costs), forming a linear function

$$J = \sum_{k=0}^{N-1} \left(\Delta T \mathbf{c}(k)^\top \mathbf{u}(k) - \mathbf{q}(k)^\top \mathbf{e}(k) \right), \quad (11)$$

which is to be maximized in the MPC formulation:

$$\max_{E, \mathbf{u}, \mathbf{e}} \sum_{k=0}^{N-1} \left(\Delta T \mathbf{c}(k)^\top \mathbf{u}(k) - \mathbf{q}(k)^\top \mathbf{e}(k) \right) \quad (12a)$$

$$\text{s.t.} \quad E(k+1) = E(k) + \Delta T \left(\eta_{\text{ch}} P_{\text{ch}}(k) - \frac{1}{\eta_{\text{dch}}} P_{\text{dch}}(k) \right), \quad (12b)$$

$$\underline{\beta} E_{\text{avail}} \leq E(k) \leq \bar{\beta} E_{\text{avail}}, \quad (12c)$$

$$E(0) = e_t, \quad \mathbf{w}(0) = w_t, \quad (12d)$$

$$\delta_{\text{ch}}(k) + \delta_{\text{dch}}(k) \leq 1, \quad (12e)$$

$$\delta_{\text{ch}}(k), \delta_{\text{dch}}(k) \in \{0, 1\}, \quad (12f)$$

$$\mathbf{u}(k) \in \mathcal{U}, \quad (12g)$$

$$\mathbf{e}(k) \in \mathbb{R}_{\geq 0}. \quad (12h)$$

The MPC framework aims to maximize the total profit (12a) while ensuring adherence to battery dynamics (12b), state constraints (12c), and initial states (12d). Exact battery model with the complementary rule is enforced with binary condition (12e). Other bounds on control actions including the limits and power balance equation are here noted as constraint set \mathcal{U} .

The optimization runs in a receding-horizon manner. At each time step, the entire sequence of optimal control actions $\mathbf{u}^*(k)$ for all $k \in \{0, \dots, N-1\}$ is computed, from which the control actions in $\mathbf{u}^*(0)$ are applied. During $\Delta T(0)$, the process is repeated with updated measurements e_t, w_t and new forecasts, ensuring continuous adaptation to changing conditions.

2.4.2. Stochastic MPC formulation

To account for the uncertainties inherent in microgrid operations, such as fluctuating renewable generation and varying load demands, we adopt a stochastic approach. This framework relies on generating multiple scenarios to represent potential realizations of uncertain variables, indexed by $s \in \mathcal{S} = \{1, \dots, N_S\}$. These scenarios allow the optimization to consider a range of outcomes, leading to decisions that are robust yet adaptive to uncertainty. The optimized variables $E(s, k)$, $\mathbf{u}(s, k)$, and $\mathbf{e}(s, k)$ may vary across scenarios, whereas parameters such as $\mathbf{p}(s, k)$, costs $\mathbf{c}(s, k)$, and weights $\mathbf{q}(s, k)$ remain the same for all scenarios.

Forecasts and scenario generation for uncertain parameters $\mathbf{w}(s, k)$ are required, but not restricted to any specific method. In this work, scenarios are generated from forecasted confidence intervals of uncertain param-

ters (e.g., PV generation, load demand). The bounds $\mathbf{w}^{\text{lb}}(k)$ and $\mathbf{w}^{\text{ub}}(k)$ at each time step k are obtained from a probabilistic forecasting model trained to produce confidence intervals at a specified percentile level, adopting the ensemble-based approach proposed by [36]. Each scenario $\mathbf{w}(s, k)$ is then constructed as a convex combination of these bounds, with an additional Gaussian perturbation to capture short-term variability:

$$\mathbf{w}(s, k) = \lambda \mathbf{w}^{\text{ub}}(k) + (1 - \lambda) \mathbf{w}^{\text{lb}}(k) + \nu(k), \quad (13)$$

where $\lambda \sim \mathcal{N}(\mu, \sigma_\lambda^2)$ is a random weight centered at $\mu = 0.5$ and scaled such that a fraction of scenarios fall outside the forecast interval, and $\nu(k) \sim \mathcal{N}(0, \sigma_\nu^2(k))$ is a zero-mean noise term estimated from historical data. Scenario generation is independent of the forecasting model itself, which only provides the point prediction and associated confidence bounds.

At the first step, non-anticipatory constraints enforce consistency across scenarios:

$$\mathbf{u}(i, 0) = \mathbf{u}(j, 0) \quad \forall i, j \in \mathcal{S}, i \neq j. \quad (14)$$

A purely robust approach that assumes the worst case leads to conservative decisions. To avoid this, we incorporate the Conditional Value at Risk (CVaR), which quantifies the expected profit in the worst-case tail of the distribution and provides a tunable level of risk aversion via the parameter $\alpha \in [0, 1]$. In our setting, CVaR_α is the expected profit over the worst α fraction of scenario outcomes, allowing a balance between risk tolerance and performance. The parameters N_S (number of scenarios) and α (risk aversion) thus offer flexibility: larger N_S improves the representation of uncertainty at higher computational cost, while smaller α values yield more conservative decisions. For implementation, we adapted the convex linear programming formulation of CVaR originally proposed in [37]. The complete derivation and the optimization model are provided in Appendix A.

The resulting stochastic MPC formulation maximizes risk-adjusted profit represented by CVaR measure while satisfying system dynamics and relaxed

constraints

$$\max_{E, \mathbf{u}, \mathbf{e}, y, z} \text{CVaR}_\alpha(y, z) \quad (15a)$$

$$\text{s.t.} \quad E(s, k+1) = E(s, k) + \Delta T \left(\eta_{\text{ch}} P_{\text{ch}}(s, k) - \frac{1}{\eta_{\text{dch}}} P_{\text{dch}}(s, k) \right), \quad (15b)$$

$$\underline{\beta} E_{\text{avail}} \leq E(s, k) \leq \bar{\beta} E_{\text{avail}}, \quad (15c)$$

$$E(s, 0) = e_t, \quad \mathbf{w}(s, 0) = w_t, \quad (15d)$$

$$\mathbf{u}(s, k) \in \mathcal{U}, \quad (15e)$$

$$\mathbf{e}(s, k) \in \mathbb{R}_{\geq 0}, \quad (15f)$$

$$z(s) \geq y - J(s), \quad (15g)$$

$$z(s) \geq 0. \quad (15h)$$

All constraints except bounds on control actions are implemented as soft to ensure the feasibility of the problem. This framework not only accounts for uncertainties but also provides a mechanism to tune the system's operation based on risk preferences and available computational resources. It ensures a balance between robustness and performance, tailored to the specific requirements of microgrid applications.

2.4.3. MPC procedure

Conventional stochastic MPC approaches without model change or scenario reduction are computationally demanding for large-scale problems. The microgrid environment that requires to respond and update power setpoints within a minute does not allow the use of the conventional stochastic MPC without any modifications.

The proposed MPC procedure combines deterministic and stochastic formulations to balance computational efficiency and robustness, addressing the challenges of solving multi-scenario optimization problems in real-time energy systems. Firstly proposed solution with two stages—stochastic and deterministic MPC—executed sequentially introduced a delay in generating the control actions. To address this limitation, the improved procedure parallelizes these stages, significantly reducing computational resources and further shortening the delays.

The principle of the two stage MPC is to reduce all scenarios to a single α -scenario s_α whose profit $J(s_\alpha)$ is equal to CVaR_α value. This enables solving the MPC as a deterministic problem with advantages of stochastic MPC, but computational time of a single scenario optimization. The stochastic

MPC serves as a pre-processing step to identify the scenario that corresponds to the desired risk level, which is then used in the deterministic MPC to compute the optimal control actions. A first step is performed at a lower frequency (e.g., every 15 minutes) as forecast updates typically do not occur with higher resolution. Once the α -scenario is determined, it remains valid for the entire interval, allowing deterministic MPC to be solved in parallel and more frequently (e.g., every minute). The full procedure consists of the following steps, which are executed in different time scales, as depicted in Figure 4:

1. **Scenario generation:** Generate N_S scenarios for uncertain parameters (e.g., PV generation, load demand) using forecasted confidence intervals. These intervals—characterized by lower and upper bounds and confidence level—are obtained from a time-series forecasting model.
2. **Stochastic MPC (infrequent):** Solve the stochastic MPC problem in form Eq. (15) using a relaxed linear model and all scenarios to determine the profit distribution across possible realizations. This model omits non-linear constraints to minimize computational time and is executed at a lower frequency.
3. **Identify the α -scenario:** For each uncertain parameter \mathbf{w} , compute the α -scenario $\mathbf{w}(s_\alpha)$ as the expected value of the worst $\alpha\%$ scenarios. This is mathematically expressed as:

$$\mathbf{w}(s_\alpha) = \frac{\sum_{s \in \mathcal{S}_\alpha} \pi(s) \mathbf{w}(s)}{\sum_{s \in \mathcal{S}_\alpha} \pi(s)}, \quad (16)$$

where $\mathcal{S}_\alpha = \{s \in \mathcal{S} \mid J(s) \leq \text{VaR}_\alpha\}$ is the set of scenarios with profits $J(s)$ below the Value at Risk (VaR_α), and $\pi(s)$ is the probability of scenario s . This ensures that the identified scenario reflects the desired risk level.

4. **Deterministic MPC (frequent):** Solve the deterministic MPC problem in form Eq. (12) using the full non-linear model and the single identified α -scenario s_α . This step is performed more frequently, taking advantage of the computational efficiency of single-scenario optimization while maintaining the nonlinearities in the model.
5. **Apply control actions:** Implement the first control action from the optimal sequence returned by deterministic MPC for the current time step.

6. **Receding horizon:** Update state measurements and forecasts, then repeat the procedure at the appropriate interval: stochastic MPC (procedure steps 1.–3.) infrequently and deterministic MPC frequently.

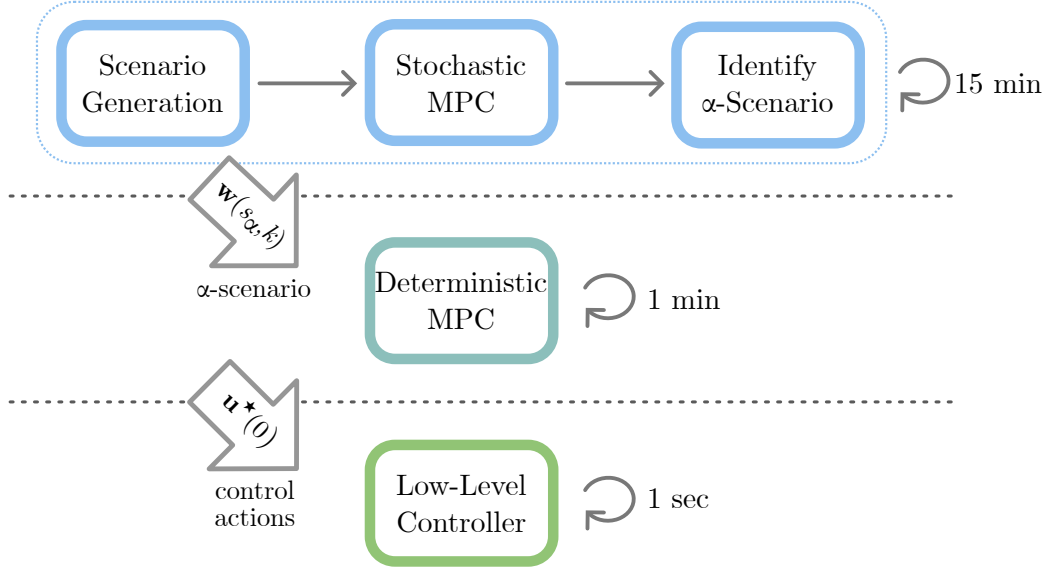


Figure 4: Major steps of the procedure in different time scales, forming hierarchical structure. Output from deterministic MPC are action commands valid for a certain time (e.g. 1 minute), serving as the setpoints for low-level controllers.

This hybrid and parallelized approach offers significant advantages. By decoupling and running stochastic and deterministic MPC stages in parallel, the procedure achieves reduced computational overhead and shorter delays between computation and implementation.

The proposed MPC framework needs a corresponding software architecture that effectively handles data flow, storage, communication, and connection with other layers of hierarchical structure. This includes managing real-time measurements and commands, storing process data such as the identified α -scenario, and coordinating the invocation of both stochastic and deterministic MPC.

2.5. From control algorithm to scalable service platform

The MPC procedure described above addresses the fundamental technical challenge of balancing computational efficiency with robustness under uncer-

tainty for a single microgrid. However, to create a commercially viable Microgrid Control as a Service, an additional dimension of complexity emerges: how to efficiently provide this control capability to multiple microgrids with diverse configurations simultaneously without substantial infrastructure investments.

Our contribution addresses this industrial scaling challenge through a novel architectural framework, enabling EMS providers to deliver sophisticated control capabilities for numerous microgrids, by systematically addressing both computational and economic dimensions of service provision.

The inherent computational characteristics of the proposed two-stage MPC procedure create opportunity for resource optimization. The sMPC component operates at 15-minute intervals with computation times approaching one minute, requiring substantial but intermittent computational resources. Conversely, the dMPC executes at one-minute intervals but completes within seconds, demanding consistent low-latency responsiveness. These distinct operational profiles directly inform our architectural design, enabling precise matching of computational resources to task requirements.

Our framework converts these computational patterns into economic advantage through three key design considerations:

- **Differentiated resource allocation:** We match computational resources to task requirements, eliminating the traditional over-provisioning approach where infrastructure must accommodate peak loads but remains idle during normal operation.
- **Dynamic scaling mechanisms:** We automatically adjust resource allocation in response to changing workloads across multiple microgrids.
- **Multi-tenancy capabilities:** We maintain strict data isolation while efficiently sharing infrastructure across customers, further improving economic efficiency.

This architectural design transforms the temporal patterns of sMPC and dMPC execution cycles into significant operational cost reductions. By eliminating idle resource time and optimizing computational efficiency, the platform achieves the economic viability necessary for commercial scale deployment. Furthermore, as the customer base expands, the architecture efficiently

supports multi-site deployment through intelligent partitioning of functionality between cloud-based and local components.

The following subsections detail our approach to mapping MPC computational tasks to appropriate infrastructure, creating a commercially viable service platform that delivers control capabilities at scale while maintaining the guarantees of the underlying MPC algorithm.

2.6. Software architecture

Building on the identified requirements, our architecture implements a hybrid approach, leveraging scalable cloud resources for centralized tasks while maintaining local control mechanisms for real-time operations.

The MaaS framework addresses use cases where customers participate in energy arbitrage and imbalance settlement by offering surplus or flexible microgrid capacity. These situations require low-latency real-time control, reliable communication, and strategic decision-making that accounts for energy market conditions, balance markets, and utility grid operator requirements. A key design principle is the separation of control layers based on their data dependency:

- **Real-time data:** The power management and control layers operate on real-time process data to maintain operational limits and respond to fast microgrid dynamics. These layers are deployed locally to ensure minimal latency, providing safe and reliable microgrid operation even under third-party faults or malicious network attacks.
- **Time-series data:** The business and energy management layers process historical and forecasted time-series data to optimize profitability and support predictive decision-making. Operating on a timescale of minutes to hours, these layers are well-suited for cloud deployment, where slightly higher latencies can be tolerated.

This separation combines the advantages of low-latency local control with the automatic scalability of cloud infrastructure, creating a framework capable of efficiently handling both rapid responses and long-term strategic planning.

To enhance memory efficiency and distribute computational load, the architecture follows a hierarchical decision-making structure from Subsection 2.2 influenced by operational timescales. At the highest level, the business management layer defines annual and monthly objectives, such as profit

maximization or carbon footprint reduction. These objectives are translated into daily to hourly energy schedules by the central energy management layer, which recomputes optimal setpoints every minute. These setpoints are then sent to the secondary control layer, which delivers commands to primary controllers responsible for power allocation at the asset level. At this stage, fast and precise execution within seconds to milliseconds is critical.

The hierarchical structure of the software architecture is designed to support scalability and promote efficient resource allocation across the infrastructure. This approach aims to reduce both upfront capital expenditures and operational expenditures, with the exact magnitude of these benefits depending on specific deployment conditions and operational parameters. The architecture’s design principles facilitate large-scale deployments through more efficient resource utilization, which leads to a reduced carbon footprint for microgrid operations.

The two-dimensional distribution of control tasks, across time-scale and resources, enables the design of a reliable and resilient architecture by combining cloud-based resources with local control hardware. By leveraging this hybrid approach, the proposed system balances centralized efficiency with localized resilience, ensuring seamless scalability across microgrids of varying complexity.

The infrastructure architecture is illustrated in Figure 5 and further detailed in the following subsections.

2.7. Cloud infrastructure

The cloud infrastructure layer ensures the availability of computational and storage resources for optimization tasks, directly influencing scalability and cost efficiency. By leveraging cloud computing, the proposed architecture achieves scalability that is effectively unbounded in terms of the number and size of operated microgrids. Cloud resources eliminate the need for CapEx on high-performance local hardware, while also providing OpEx flexibility for variable workloads. In contrast, local deployments require over-provisioning to handle peak loads, leading to resource inefficiency. Additional benefits of cloud-based infrastructure include high availability, uptime guarantees, resilience, and security, which are essential for safety-critical applications.

Typically cloud providers, including Alibaba, AWS, Azure, Google, IBM, Oracle, and Tencent, offer a range of computing services. While these are broadly categorized as cloud compute services, varying degrees of flexibility

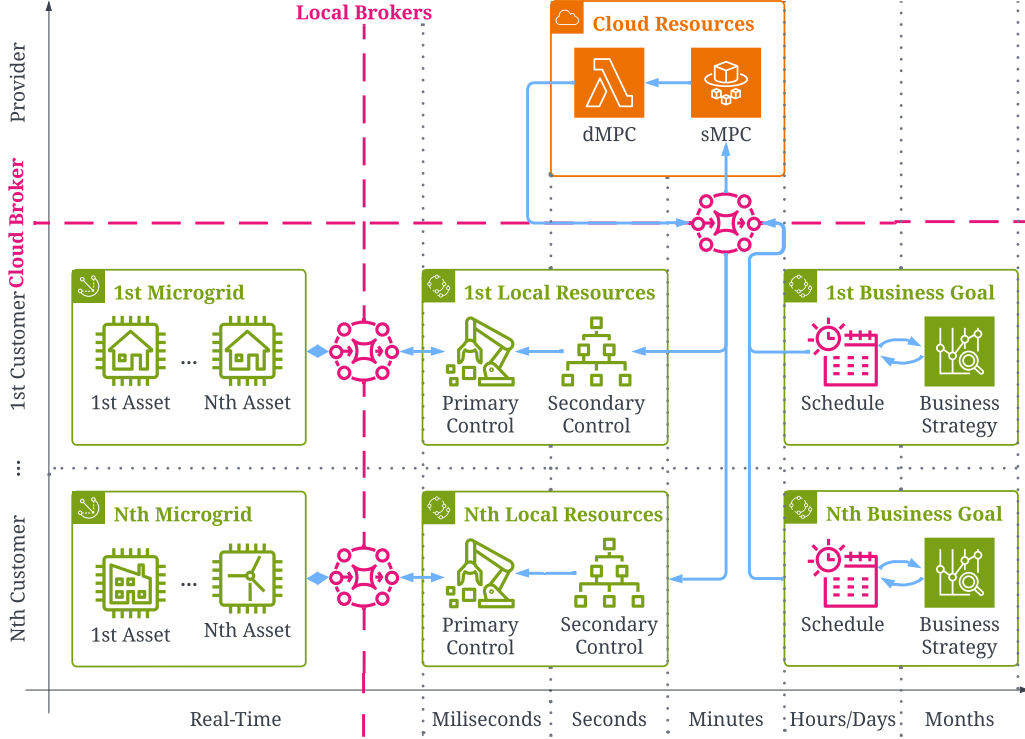


Figure 5: Individual customers operate their microgrids with varying number of assets using local control units, while the MaaS provider offers cloud-based computation to reduce CapEx and optimize OpEx based on the customers business goals.

and control have led to the emergence of three categories of solutions: Function as a Service (FaaS), Container as a Service (CaaS), and Infrastructure as a Service (IaaS). Each option presents trade-offs in terms of OpEx, and operational complexity. A comparison of key features and billing models of these services is provided in Table 2, illustrating the trade-offs between pricing, flexibility, and complexity.

One of the primary cost advantages of cloud-based services is the pay-as-you-go billing model, which eliminates CapEx and idle-time costs by ensuring that resources are billed only when in use. The selection of an appropriate cloud service depends largely on task duration, frequency, and intensity, as well as the required levels of availability and scalability.

For instance, FaaS, despite its high per-unit price, offers the lowest cost

Table 2: Comparison of cloud compute solutions.

Feature	FaaS	CaaS	IaaS
Provisioning	Serverless	Serverless	Managed
Payment Model	As-you-go	As-you-go	Per-instance
Billing Period	Millisecond	Second	Hour
Billing Unit	GB-s + re-requests	(vCPU + GB)-h	instance-h
Pay for Idle Time	×	×	✓
Price per Unit	High	Medium	Low
Price per Request	Low	Medium	High
Task Duration	Short	Medium	Long
Flexible Scaling	✓	✓	×
Reservation Time	Millisecond	Minute	Hour
Startup Time	Milliseconds	Seconds	Minutes
Flexible Setup	Environment	Container	Virtual Machine
Maintenance	None	Low	Medium

for short and frequent tasks, with OpEx computed as:

$$\mathcal{C}_{\text{FaaS}} = \sum_{r=1}^R \left(c_r + \sum_{t=1}^T (m c_m(r)) \right), \quad (17)$$

where R is the number of requests, c_r is the request cost, T is the execution time, m is the allocated memory, and $c_m(r)$ is the memory cost.

In contrast, CaaS is better suited for longer, less frequent tasks due to its higher startup time and one-minute minimum duration. Its OpEx is given by:

$$\mathcal{C}_{\text{CaaS}} = \sum_{r=1}^R \sum_{t=1}^T (v c_v + m c_m), \quad (18)$$

where v is the number of vCPUs, c_v is the vCPU cost, m is the allocated memory, and c_m is the memory cost.

For long, resource-intensive tasks, IaaS instances are more suitable despite their higher cost. The OpEx for IaaS is:

$$\mathcal{C}_{\text{IaaS}} = \sum_{t=1}^T (i(t, r) c_i), \quad (19)$$

where c_i is the instance cost and the number of instances i is determined by the task duration and the number of requests:

$$i = \frac{r \cdot t}{\mathcal{R}_{\text{conc}} t_i}, \quad (20)$$

with $\mathcal{R}_{\text{conc}} = \min\left(\frac{i_v}{v}, \frac{i_m}{m}\right)$ representing the number of concurrent requests and t_i the active instance time.

Figure 6 illustrates influence of request duration on monthly costs, associated with individual services, executing requests every minute over 30 days, which is equivalent to 43 200 requests, requiring 0.5 vCPU and 2048 MiB of memory. The figure highlights that each service has an optimal use case. While FaaS is the most cost-effective for short tasks under 12 seconds, IaaS becomes more efficient for tasks up to 60 seconds. For tasks longer than 60 seconds, CaaS provides finer scalability and potential to reduce OpEx. This advantage arises because IaaS requires entire instances to be added or replaced, with consideration on OS overhead consuming from 5% to 30% of resources, whereas CaaS scales resources with greater precision, allowing cost savings.

Examining sensitivity of the monthly costs to the request duration, Figure 7 shows that the cost of FaaS is most sensitive to the change in parameters due to its high per-request price and lowest threshold on resource allocation. While FaaS enables significant savings, if chosen for inappropriate tasks, it can lead to significant cost overruns. Meanwhile, CaaS and IaaS are less sensitive to changes in request duration, with CaaS offering the lowest price for intensive tasks. IaaS has the lowest sensitivity, making it suitable for reservation of resources for fluctuating workloads. All cost calculations are based on AWS pricing data as of March 2025, using the cost models defined in Equations (17), (18), and (19). These calculations reflect realistic hypothetical workloads that would emerge from varying the number of scenarios in MPC optimization, computation frequency, and the number of microgrids under management—parameters directly relevant to our EMS deployment described in the case study.

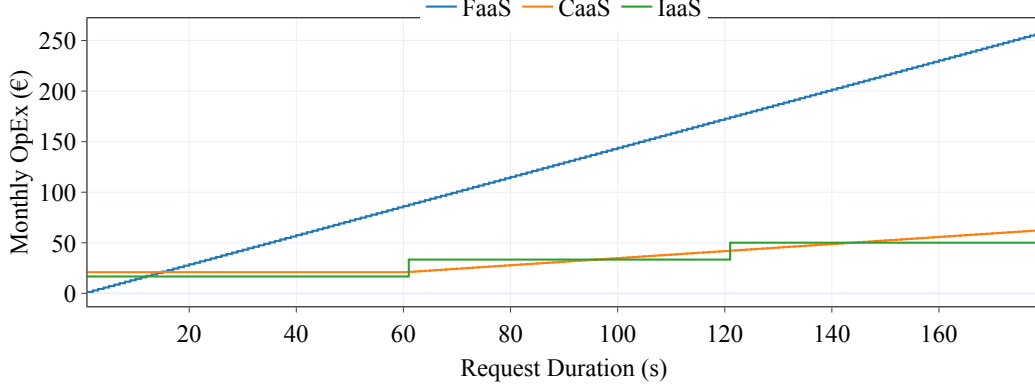


Figure 6: Monthly OpEx for different request durations requiring 2048 MiB of memory and 0.5 vCPU. FaaS has the lowest OpEx for short requests but grows rapidly. OpEx for CaaS is constant until request durations of 60 seconds, then grows steadily. IaaS grows every time additional instance is activated.

Based on our analysis of service characteristics, we have implemented our MPC framework using a hybrid deployment strategy. In our previous implementation, the MPC framework was deployed on reserved IaaS instances, which introduced operational challenges including persistent idle-time costs, over-provisioning, and complex manual scaling requirements [35]. Our current architecture distributes MPC tasks across two serverless services based on their computational profiles: sMPC tasks (running at 15-minute intervals with execution times around one minute) are assigned to CaaS, leveraging its capacity for longer-duration batch processing, while dMPC tasks (executing every minute and completing within seconds) are deployed on FaaS, benefiting from millisecond-level billing and eliminating idle charges. This mapping of tasks to appropriate services optimizes both computational performance and operational costs.

2.8. Local control units

Local control units form the secondary control layer, operating with second-level response times. They maintain microgrid stability through rapid state adjustments, even during communication failures with the central EMS. This distributed control approach enhances system resilience and reliability.

Industrial-grade microcontrollers (PLCs or industrial PCs) handle these control tasks with minimal hardware requirements. During communication

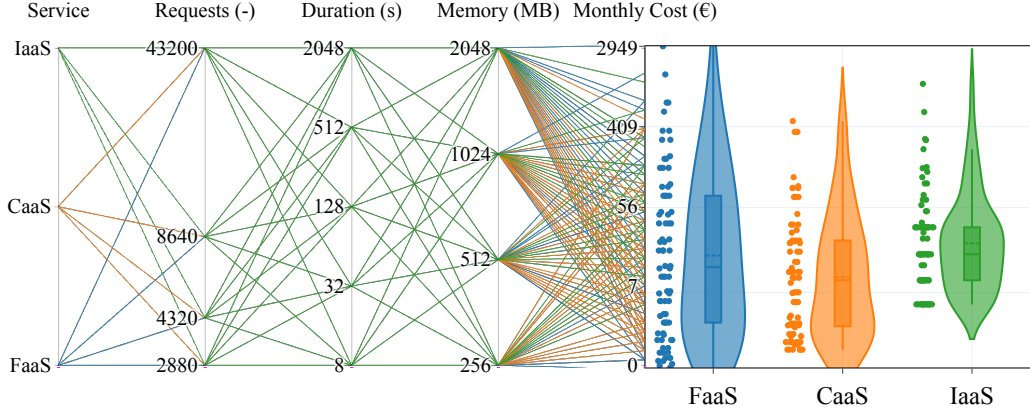


Figure 7: Influence of parameter variations on monthly cost of different services based on AWS pricing (March 2025). FaaS offers potential for saving if selected for proper case, while CaaS has lowest cost for high workloads. Cost of IaaS is the most stable under various workloads.

failures, local controllers can independently optimize power dispatch and regulate loads using locally stored data. This architecture reduces central system overhead while ensuring uninterrupted microgrid operation.

The primary control layer, operating at millisecond resolution, is implemented directly in power electronic converters using PID control. Further details of this layer are not covered within the paper. Together, these layers ensure precise power reference tracking and system stability.

3. Case study

To prove the effectiveness of the proposed EMS framework complemented with the architecture, we present a study from the actual operation realized in cooperation with our business partner. Among multiple commercial and industrial EMS applications, a suitable microgrid is selected for the case study to demonstrate the results. A detailed description of the microgrid system parameters, objectives, environment, control setup, and operational outcomes is provided.

3.1. Microgrid description

The microgrid is located in the Netherlands and consists of a commercial building that relies on the EMS to control entrusted energy assets and manage its energy supply efficiently. The microgrid system is equipped with a 220

kWp rooftop photovoltaic system, a battery energy storage system with the rated power and capacity 180 kW/558 kWh, electric vehicle chargers of size 200 kW that are considered as controllable load, and the building’s general energy demand belonging to the non-controllable load.

The microgrid maintains a physical connection to the Dutch distribution grid, subject to predefined import and export power limits. Energy transactions are structured through a third-party energy supplier, responsible for purchasing and selling electricity on the day-ahead market (DAM). This supplier also provides the day-ahead consumption and generation schedule, referred to as the day-ahead diagram (P_{DA}). Any deviations from this schedule are settled based on the real-time market imbalance prices.

3.2. Objectives

At the uppermost level, the customer defines the overarching objectives of the EMS—prioritizing profit generation while ensuring a reliable microgrid operation. These objectives are derived from the environment in which the microgrid is set and the actual billing model, which accounts for dynamic energy day-ahead market prices and additional fees that forms the spread between buying and selling electricity price.

Among the cost-related factors under the EMS control, imbalance settlement has the most significant financial impact. Deviations from the day-ahead schedule result in either penalties or profits based on the real-time market conditions. The EMS actively controls battery storage and other power-flexible assets to strategically participate in imbalance management.

Additionally, the EMS aims to maximize the self-consumption of PV generation while dynamically curtailing excess production when market conditions make grid exports unprofitable. EV charger power availability is also controlled with respect to its operational demands. The EMS continuously evaluates price signals, forecasts, and system states to determine optimal times for curtailment and load adjustments, ensuring cost-effective and reliable operation. Imposed grid export and import limits determine the need for the active peak-shaving functionality, as the limit exceeding is associated with a financial penalty.

Besides the objectives, the business layer also decides about the risk aversion that it is willing to handle, which is reflected in the α parameter of the CVaR measure.

3.3. Parametrization

The prediction horizon of the model predictive controller was set to be 24 hours because a daily pattern occurs in power profiles. The sampling time ΔT was set to 15 minutes, which is the usual billing period, then $N = 96$. From the microgrid composition, it results that we need to forecast PV generation and load demand that form the uncertainty array \mathbf{w} . The probabilistic forecasts were obtained using a recurrent neural network (RNN) with long short-term memory (LSTM) architecture. The model was trained on 30 days of historical measurements to output point forecasts and 90% confidence intervals at a 15-minute resolution. These intervals were then used as inputs to the scenario generation procedure described in Section 2.4.2. Energy price predictions of day-ahead market for the relevant horizon are known with certainty, so they are part of the cost array \mathbf{c} . Similarly are treated the imbalance prices, the values for the current time step are being published by the grid operator TenneT, and for the rest the estimation is assumed.

The control variables in \mathbf{u} to be computed include the battery power flow, PV power curtailment, controllable load of EV chargers, and power exchange with the main grid. The battery power flow is constrained by its rated power, given in the parameter array \mathbf{p} . The controllable EV charging load and PV power are bounded by actual demand. Power exchange with the main grid is subject to fixed import and export limits set by the grid operator. The battery energy level E is also a decision variable, but it is not directly controlled as it depends on power decisions. Each action is associated with a cost: battery operation account for degradation, PV curtailment represents foregone earnings from potential subsidies, and reducing EV charging power reflects the trade-off with user convenience due to extended charging times.

The particular instance of the microgrid is characterized by the following variables in array form:

$$\mathbf{u} = [P_{\text{ch}}, P_{\text{dch}}, P_{\text{imp}}, P_{\text{exp}}, P_{\text{short}}, P_{\text{long}}, P_{\text{curt}}, P_{\text{adj}}]^\top, \quad (21a)$$

$$\mathbf{w} = [P_{\text{RES}}, P_{\text{load}}]^\top, \quad (21b)$$

$$\mathbf{p} = [E_{\text{nom}}, \underline{\beta}, \bar{\beta}, \eta_{\text{ch}}, \eta_{\text{dch}}, P_{\text{DA}}, \bar{P}_{\text{ch}}, \bar{P}_{\text{dch}}, \bar{P}_{\text{imp}}, \bar{P}_{\text{exp}}]^\top, \quad (21c)$$

$$\mathbf{c} = [-c_{\text{ch}}, -c_{\text{dch}}, -c_{\text{imp}}, c_{\text{exp}}, -c_{\text{short}}, c_{\text{long}}, -c_{\text{curt}}, -c_{\text{adj}}]^\top, \quad (21d)$$

$$\mathbf{e} = [\underline{c}_{\text{E}}, \bar{c}_{\text{E}}, \epsilon_{\text{imp}}, \epsilon_{\text{exp}}]^\top, \quad (21e)$$

$$\mathbf{q} = [\underline{q}_{\text{E}}, \bar{q}_{\text{E}}, q_{\text{imp}}, q_{\text{exp}}]^\top. \quad (21f)$$

Values of parameters which are fixed for all steps k are given in Table 3, a fixed fee is stated for c_{imp} and c_{exp} , to which variable price from day-ahead market is added.

Table 3: Microgrid parameters and their values.

Parameter description		Symbol	Value
Storage state of health	(%)	S	99.6
Nominal energy capacity	(kWh)	E_{nom}	558.0
Available energy capacity	(kWh)	E_{avail}	555.8
Min. state of charge	(%)	β	10.0
Max. state of charge	(%)	$\bar{\beta}$	95.0
Charging efficiency	(%)	η_{ch}	95.0
Discharging efficiency	(%)	η_{dch}	95.0
Max. charging power	(kW)	\bar{P}_{ch}	180.0
Max. discharging power	(kW)	\bar{P}_{dch}	180.0
Max. import power	(kW)	\bar{P}_{imp}	45.0
Max. export power	(kW)	\bar{P}_{exp}	150.0
Cost of imported energy	(€ / MWh)	c_{imp}	100.0
Cost of exported energy	(€ / MWh)	c_{exp}	1.5
Cost of charging	(€ / MWh)	c_{ch}	0.0
Cost of discharging	(€ / MWh)	c_{dch}	25.0
Cost of RES curtailment	(€ / MWh)	c_{curt}	5.0
Cost of load adjustment	(€ / MWh)	c_{adj}	600.0
Penalty for lower SoC limit	(€ / MWh)	\underline{q}_{E}	1×10^6
Penalty for upper SoC limit	(€ / MWh)	\bar{q}_{E}	1×10^6
Penalty for import power limit	(€ / kW)	q_{imp}	1×10^5
Penalty for export power limit	(€ / kW)	q_{exp}	1×10^5

The parameters that are time-varying are not shown here, but analysis is provided for imbalance prices, as they have a significant impact on overall control behavior. Figure 8 shows the distribution of imbalance settlement prices over one month, September 2024. The lowest price for both, long

and short, was -1000.0 €/MWh, while the highest was 1850.0 €/MWh, the median price was 67.0 €/MWh for long and 79.2 €/MWh for short. When both prices are positive, the microgrid is penalized for being in shortage and remunerated otherwise, when both prices are negative, the microgrid is penalized when being in energy surplus, and when prices differ in sign, the microgrid is penalized for any imbalance. In any 15-minute period, both prices are either equal or the long price is greater.

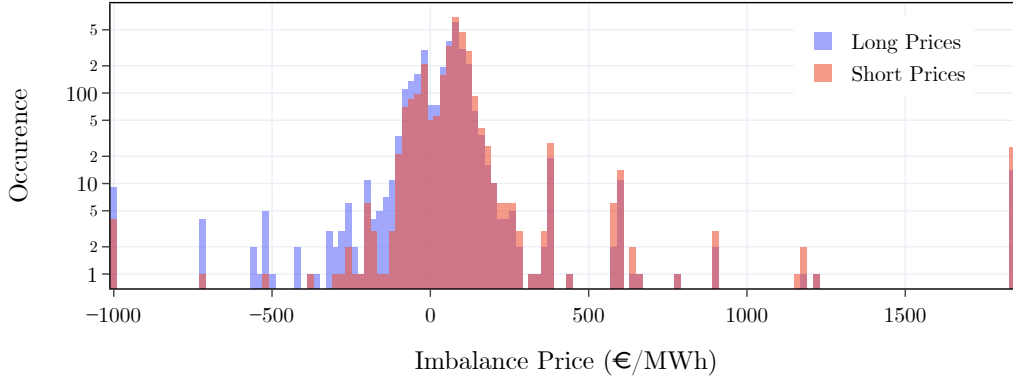


Figure 8: Distribution of imbalance prices, positive prices occur more often, several significant occurrences of high prices.

3.4. Risk aversion

The level of risk aversion depends on operator preferences and can be tuned through the parameter α . To quantify its impact and provide practical guidance for its selection, we conducted a backtest simulation with historical measurements and forecasts. Control actions were applied in a closed-loop simulation and total profits were calculated for one day. To account for uncertainty, the Monte Carlo method generated 100 realizations of stochastic parameters, with results evaluated across a range of α values (0–100% in 25% increments).

High values of α correspond to more aggressive, risk-seeking control: the EMS assumes nominal forecast realizations, allowing greater use of battery

capacity for revenue-generating actions but exposing the system to higher potential losses. Conversely, low values of α yield conservative decisions: the EMS prepares for adverse scenarios, reducing risk but also limiting profit opportunities. For instance, in a microgrid with tight grid limits, a conservative controller may keep extra energy in storage to cover unexpected peaks, even at the cost of charging during high prices.

The results are summarized in Table 4, which shows profit distributions across percentiles. At $\alpha = 0\%$, the controller is most robust and performs best under adverse realizations, while at $\alpha = 100\%$, it achieves higher profits under normal conditions but suffers much larger losses in unfavorable cases. The median case illustrates typical performance, while the worst case shows the unavoidable low-profit outliers that arise under unfavorable market or weather conditions.

Table 4: Total profit/loss per day from closed-loop simulation, 100 runs for each α level. In bold is marked the best choice of α parameter for each percentile. Parameter α being 0% represents the most robust controller, while $\alpha = 100\%$ the conventional.

α	(%)	0	25	50	75	100
Best case	(€)	45	45	47	54	55
25% quartile	(€)	22	23	24	30	6
Median	(€)	8	10	11	14	-37
75% quartile	(€)	-1	2	3	-40	-89
Worst case	(€)	-163	-171	-204	-336	-337

The influence of α on scenario selection is further illustrated in Figure 9, which shows forecast intervals, generated scenarios, and the selected trajectory under different α values. For higher α , the selected scenario is closer to the median, while lower α shifts toward the worst-case realization. This behavior varies dynamically with market conditions, limits, and renewable production.

In practical terms, operators can choose α based on the maximum loss they are willing to tolerate, as indicated by the percentile outcomes in Table 4. For example, at $\alpha = 25\%$ the worst-case loss is limited to -171€, while median and quartile returns remain positive, which represents a balanced compromise between robustness and profitability.

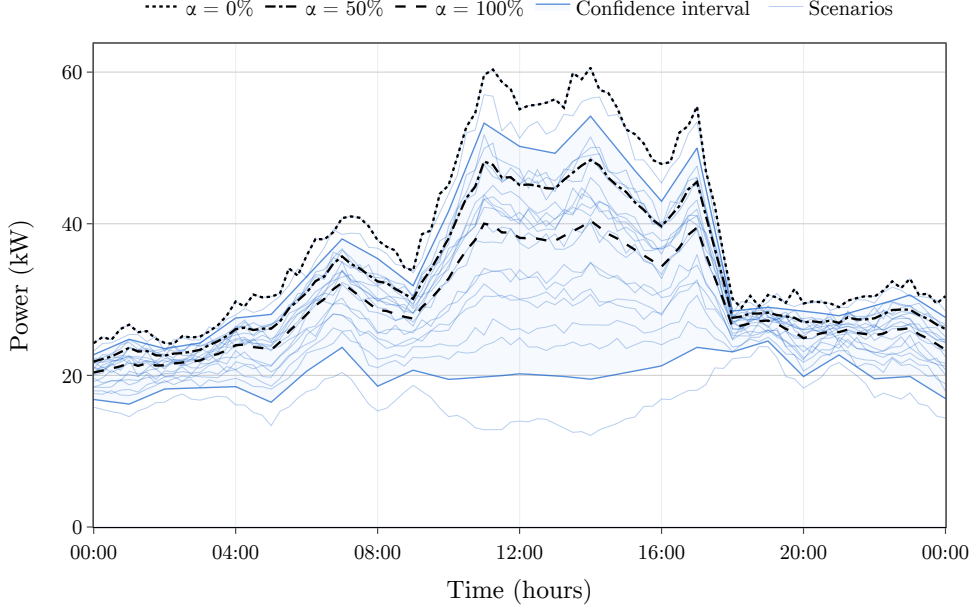


Figure 9: Selected α -scenario of consumption demand P_{cons} as the expected value of the worst $\alpha\%$ scenario realizations for different α levels. Scenarios are generated from known 90% forecasted confidence interval, 20 of them are visualized.

3.5. Operation performance

The EMS has been deployed in commercial operation for more than three months (starting in September 2024), of which an example day (8th of October 2024) is showcased in Figure 10. The day was chosen as representative of typical operation, with fluctuations in PV generation and load demand. The EMS effectively managed the microgrid, optimizing battery usage, PV curtailment, and EV charging to maximize profit while adhering to operational constraints. Figure 10 compares the realized grid exchange under EMS control with a baseline case without EMS. Since real-world experiments with alternative control strategies were not feasible, the “without EMS” case was estimated by removing all EMS control actions and recalculating the power flows using the power balance equation under the same demand and generation profiles. Economic results for the baseline were then computed from these recalculated power profiles using the same price signals, ensuring a consistent counterfactual for assessing the cost efficiency of the EMS.

The power profiles in Figure 10 are 15-minute averages. It can be seen, in several intervals the power demand was higher than combined self-production and grid import limit. Thanks to the storage power supply, the limits were not exceeded and demand could be satisfied. The revenue/cost decomposition to multiple revenue streams in hourly intervals is visualized in Figure 11. Difference between revenues and costs is the profit (or loss if value is negative).

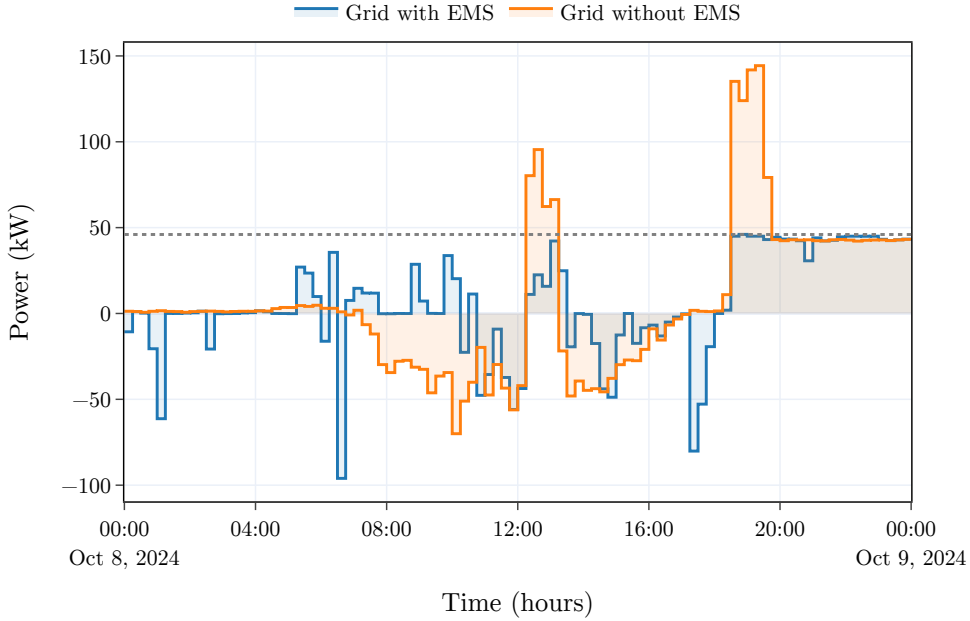


Figure 10: Example day of microgrid operation, figure compares power grid exchange with (realized) and without (estimated) applied EMS control. Dotted line represents the import limit.

Figure 12 visualize one month of microgrid operation using a heatmap representation, allowing temporal patterns to be observed at a glance. The microgrid effectively manages its energy resources, maximizing self-consumption of PV generation, while using grid exchanges to benefit from imbalance price opportunities. The third panel presents the battery power flow: charging occurs during midday PV surplus and low-price intervals, while discharging aligns with demand peaks and high-price periods.

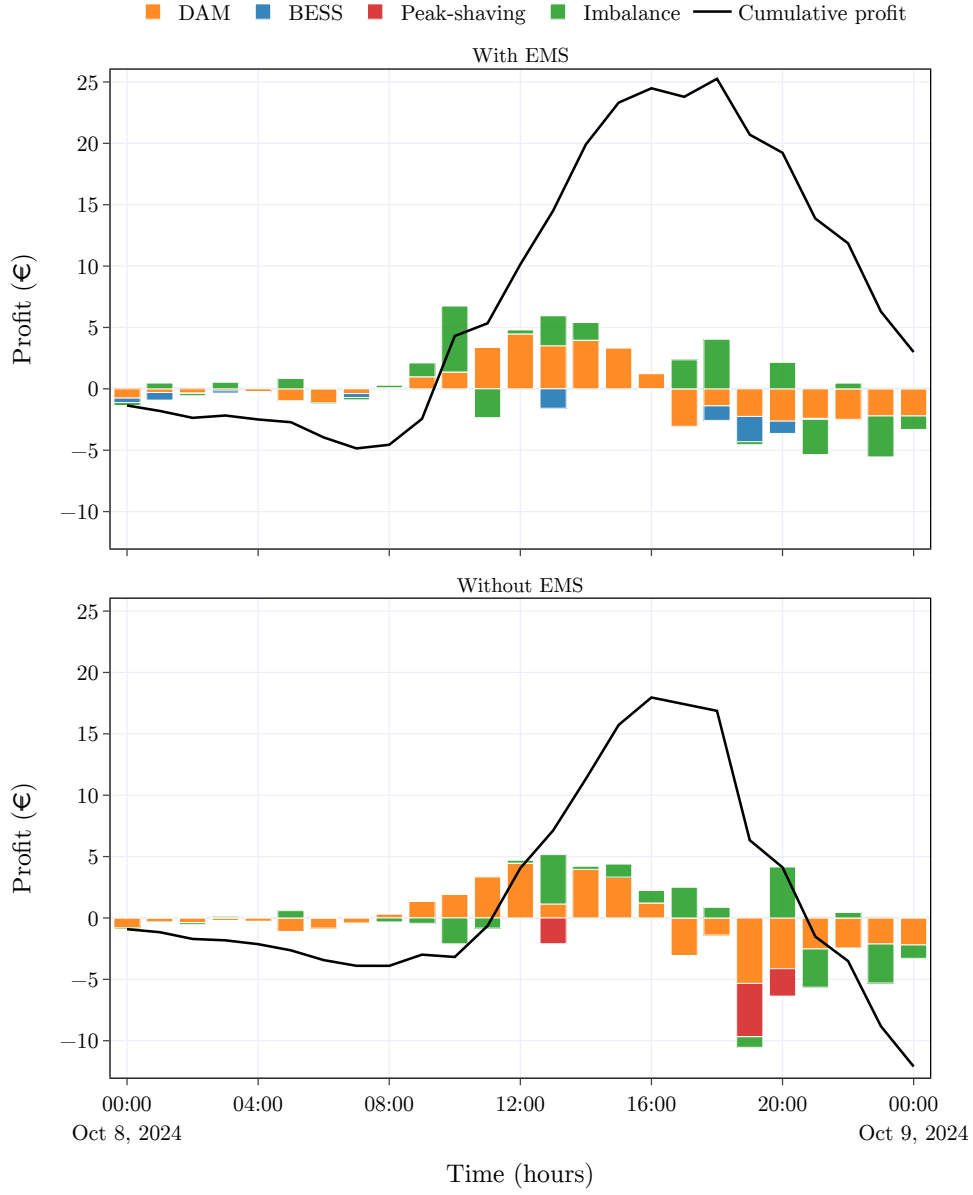


Figure 11: Revenue/cost decomposition of operating the microgrid with and without EMS control actions. Total profit/loss (revenues minus costs) for example day was 3.0 €, compared to -12.1 € without predictive control, which made microgrid operation profitable.

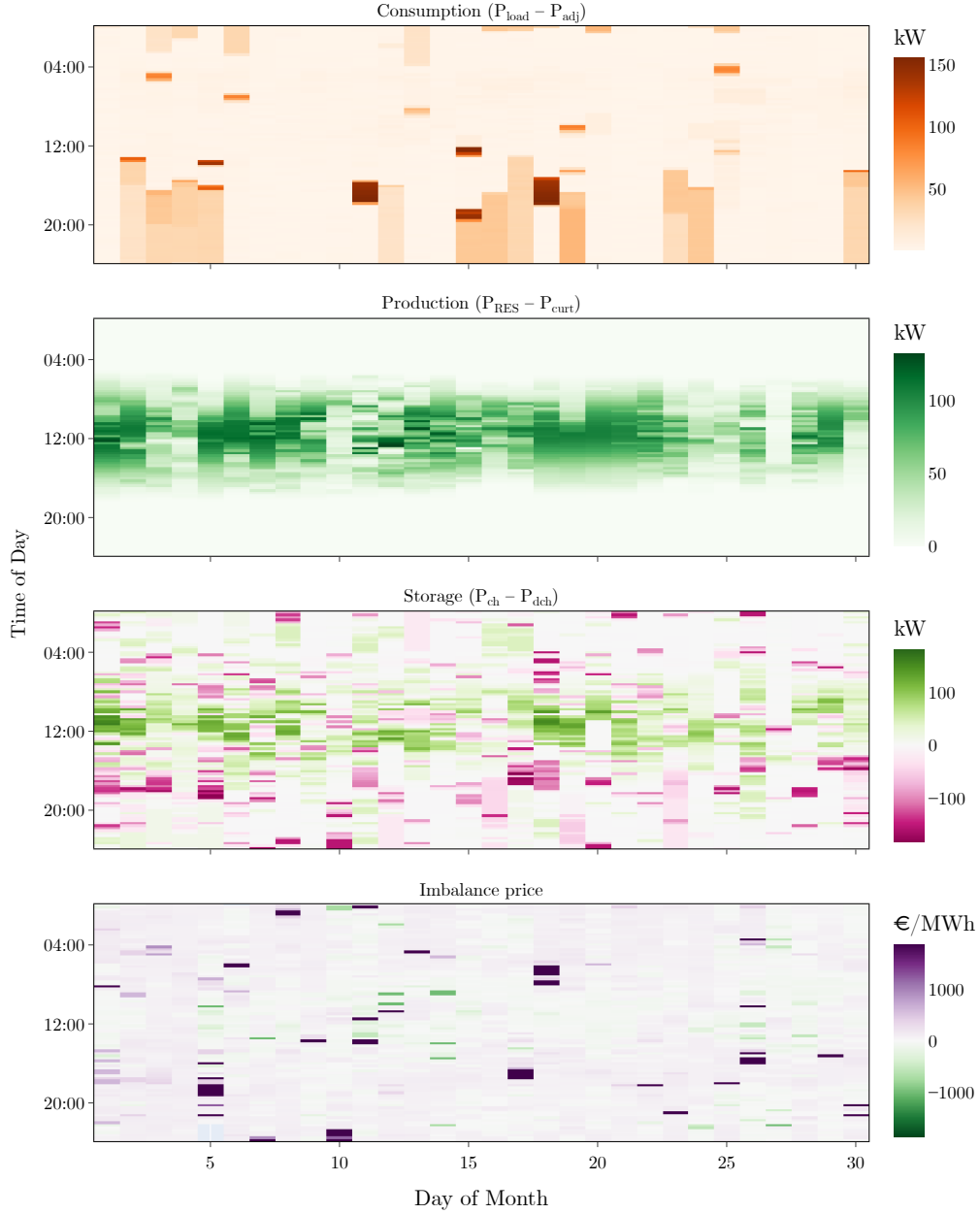


Figure 12: Heatmap visualization of microgrid operation over one month: (a) consumption, (b) PV production, (c) storage power flow (positive is charging, negative is discharging), and (d) imbalance price. The representation reveals recurring daily patterns and shows how battery actions are simultaneously aligned with all the local power variations and market signals.

The EMS utilizes flexibility of assets, predominantly the flexibility of battery storage, to maximize operating profit. Operating profit add up energy volumes bought and sold on the day-ahead market, imbalance settlement costs, battery degradation costs, and peak-shaving costs, all stated separately in Table 5. Negative cost value for day-ahead market indicates more energy needed to be imported rather than exported, due to PV system being sized smaller than microgrid energy demands. The most significant cost component directly influenced by the EMS control actions is the imbalance settlement. The optimal control of microgrid assets turned imbalance cost of -600.4 € into revenue of 2471.1 € over the month period. Peak-shaving costs are associated with the energy (in amount of 444.3 kWh during the examined period), which without the optimal control would not be supplied, or grid import limit would be exceeded. The EMS actions reduced the peak-shaving costs to zero, as the battery was used to supply the excess demand. Visualization of the cost components for example day in Figure 11.

Table 5: Comparison of operating revenues/costs over the month period with and without EMS actions. Positive values denote revenues, negative values denote costs.

Revenue/cost type		Without EMS	With EMS
Imbalance	(€)	-600	2471
Day-ahead market	(€)	-326	-405
Battery usage	(€)	0	-210
Peak-shaving	(€)	-267	0
Total profit	(€)	-1193	1856

The EMS yield is evaluated as the difference between the profit with and without EMS. For benchmarking purposes, the total yield is related to the unit of battery-rated power. Results are summarized in Table 6, with the achieved annual revenue of 159 €/kW exceeding investors' expectations (according to the study in [38], an annual revenue of 145 €/kW ensures a 10% Internal Rate of Return). Benchmarking against existing operational strategies is difficult, as the results are highly dependent on market conditions. A similar framework developed in [9] demonstrated its effectiveness in the modeled Belgian balancing market, achieving revenue of 5.3 €/kW/month and 21.8 €/kW/month for a month with low (33 €/MWh) and

high (216 €/MWh) average imbalance prices, respectively. For comparison, we achieved 17 €/kW/month for the month with an average imbalance price of 67–79 €/MWh.

Table 6: Yield of operating the battery storage over the month and year period.

		1 Month	1 Year [*]
Total yield	(€)	3049	28 700
Yield per kW	(€/kW)	17	159

^{*} Value extrapolated from 3 months of operation.

The transferability of the achieved revenues to other microgrids is not straightforward. The provided results are valid for a 0.33 C-rate (3-hour discharge) storage and given composition, but long-term operation of other microgrids within the same markets supports similar revenues.

To illustrate the versatility and scalability of developed solution, Table 7 provides an overview of heterogeneous microgrids where the EMS was deployed. Developed EMS wrapped into Microgrid Control as a Service supports various microgrid configurations and applications to be easily adopted and scaled.

3.6. Computational benchmarking of MPC formulations

To evaluate the computational efficiency of the proposed two-stage MPC, we benchmarked it against alternatives. The main classes of formulations are:

- **Deterministic MPC**, equivalent to stochastic MPC with $N_S=1$ full model scenario.
- **Full stochastic MPC**, this refers to conventional stochastic MPC, where the nonlinear full model is solved simultaneously across N_S scenarios. This provides the most rigorous risk-aware solution but is computationally prohibitive beyond very small N_S .
- **Relaxed stochastic MPC**, where nonlinearities are simplified to obtain approximate solutions across multiple scenarios. This approach reduces runtimes but sacrifices model fidelity.

Table 7: Overview of microgrids, their composition, and functionalities to showcase support for wide range of uses. Presented microgrid is in the first row.

Facility	Composition	Functionality
Commercial building	PV (220 kWp), BESS (558 kWh), EV chargers (200 kW)	Imbalance optimization, Peak-shaving
Commercial building	PV (150 kWp), BESS (150 kWh), EV chargers (132 kW)	Price arbitrage, Peak-shaving
Industrial park	PV (600 kWp), BESS (2×150 kWh), Heat pumps	Imbalance optimization, Peak-shaving
Industrial park	PV (280 kWp), BESS (150 kWh), Heat pumps	Imbalance optimization
Industrial park	PV (350 kWp), BESS (150 kWh)	Price arbitrage, Self-sustainability
Industrial park	PV (500 kWp), BESS (2×150 kWh)	Price arbitrage, Peak-shaving
Mountain resort	PV (800 MW), BESS (558 kWh)	Peak-shaving, Self-sustainability

- **Proposed two-stage MPC**, which first applies relaxed stochastic MPC to identify the α -scenario and then solves deterministic MPC with the full nonlinear model. This procedure preserves model fidelity while scaling to larger N_S .

The robust MPC (worst-case) can be obtained within this framework by setting $\alpha = 0\%$, which does not materially affect computational effort and is therefore not listed separately.

Benchmarking was performed on the microgrid defined in Section 3.3, using a MacBook M2 (8-core CPU, 24 GB RAM) and GLPK solver. Average and maximum runtimes over 50 runs are reported in Table 8. As expected, the runtime of full stochastic MPC grows rapidly with N_S , reaching solver

timeouts beyond very few scenarios. The relaxed formulation remains feasible at larger N_S , though with increasing runtime. The proposed two-stage method inherits the runtime of the relaxed formulation for identifying the α -scenario, while the deterministic stage has fixed runtime independent of N_S , equal to that of $N_S=1$.

Table 8: Computation time vs. number of scenarios N_S . Results averaged over 50 runs on the same hardware/solver. † denotes timeout after 60 minutes. Note: $N_S=1$ corresponds to deterministic MPC.

N_S	Full sMPC	Relaxed sMPC
	Max/Avg (s)	Max/Avg (s)
1	3.0 / 2.3	2.3 / 2.0
5	365.8 / 292.0	4.7 / 4.4
10	1813.8 / 1684.4	11.1 / 9.7
25	†/ †	52.0 / 48.7
50	†/ †	193.4 / 182.7

The two-stage MPC combines relaxed sMPC (Stage 1) with deterministic MPC (Stage 2) executed in parallel, so the overall runtime is not the simple sum of both stages, but close to deterministic MPC while the robustness of scenario-based formulations is retained. Higher values of N_S improves resolution of uncertainty representation, but with diminishing returns beyond a certain point. A rigorous procedure for selecting the optimal number of scenarios is outside the scope of this study and may be left for future work.

3.7. Cost analysis of cloud deployment for the MaaS provider

This analysis evaluates the computational OpEx incurred by an EMS provider when operating the proposed EMS using a hybrid cloud deployment of FaaS and CaaS. The simulation of the MaaS provisioning to microgrid located in Netherlands is used to illustrate how execution frequency, computational requirements, and cloud service billing models affect the overall OpEx for the provider.

3.7.1. Computational performance

The provider needs to allocate execution of two MPC tasks—sMPC and dMPC—as detailed in Subsection 2.4.3 and illustrated in Figure 4. To

illustrate the economic viability of the proposed cloud deployment strategy, we conducted 100 benchmark simulations using the actual configuration parameters from our case study installation. Table 9 summarizes the key performance metrics from these benchmarks.

Table 9: Performance summary of sMPC (30 scenarios) and dMPC workloads. Execution frequency is the time between two consecutive executions of the task, while execution time is the time needed to compute the solution.

Metric	sMPC	dMPC
Max execution time (s)	98	3.0
Avg execution time (s)	83	2.4
Execution frequency (s)	900	60
Max memory used (MB)	736	180
Avg memory used (MB)	702	175

The computational performance metrics were achieved using GLPK, an open-source solver released under the GNU General Public License, which is integrated into our EMS framework. This choice of non-commercial solver is critical for maintaining sustainable operational costs, as proprietary alternatives would introduce prohibitive licensing expenses. While open-source solvers typically demand more computational time, our hybrid cloud deployment strategy specifically accounts for this limitation by ensuring that the dMPC execution time remains within the one-minute threshold, making the trade-off between computational efficiency and cost-effectiveness work to our advantage.

3.7.2. Cost breakdown

Using the equations presented in Subsection 2.7, we calculated the operational costs based on actual request frequencies, execution times, and memory requirements derived from our benchmark simulations. The monthly OpEx analysis considers 30 days of continuous operation for the installations in our case study. The dMPC task executes every minute (43 200 monthly executions), while the sMPC runs at 15-minute intervals (2880 monthly executions). Table 10 presents a comparative analysis of the total monthly OpEx for both control tasks across different cloud service models.

Table 10: Monthly OpEx for EMS control tasks.

service	dMPC	sMPC	dMPC/req.	sMPC/req.
FaaS (€)	0.38	3.38	0.000 01	0.0012
CaaS (€)	10.49	1.14	0.000 24	0.0004
IaaS (€)	3.02	8.35	0.000 07	0.0029

The results indicate that deploying dMPC on FaaS is the most cost-efficient approach, as it benefits from millisecond-level billing, making it ideal for short-duration tasks. Meanwhile, sMPC is best suited for CaaS, where costs scale linearly with memory usage and execution time, subject to a minimum billing period of one minute. While the cost difference between CaaS and FaaS may appear marginal in this specific case study, this is primarily due to the relatively low memory requirements of this single-microgrid illustrative implementation. At larger scales, where higher memory requirements are inevitable, the cost advantage of CaaS would become even more pronounced, as seen for request durations of 60 second in Figure 6. The hybrid deployment strategy eliminates idle-time costs and enables dynamic resource scaling, avoiding the higher expenses associated with IaaS or dedicated local infrastructure. Importantly, the additional OpEx introduced by the EMS operation remains three orders of magnitude lower than the revenue it generates (see Table 5), confirming the economic viability of the proposed architecture for the MaaS provider.

To further demonstrate why deployed of sMPC on CaaS is scaling-proof, we conducted sensitivity analysis in Figure 13, depicting the cost sensitivity to scaling to the number of scenarios in the optimization problem. The results show that for low number of scenarios, FaaS deployment is more viable. Nevertheless, increasing number of scenarios as part of the risk mitigation strategy, enhances customer profitability. Above 25 scenarios, the OpEx for sMPC on CaaS remains the lowest, even as the number of scenarios increases. In contrast, FaaS costs rise exponentially with the number of scenarios, making it less cost-effective for large-scale deployments.

This analysis confirms that a hybrid FaaS-CaaS deployment offers a cost-efficient EMS architecture while ensuring computational reliability. The dynamic allocation of tasks minimizes cost overhead, making the approach highly scalable for large-scale microgrid deployments with diverse assets.

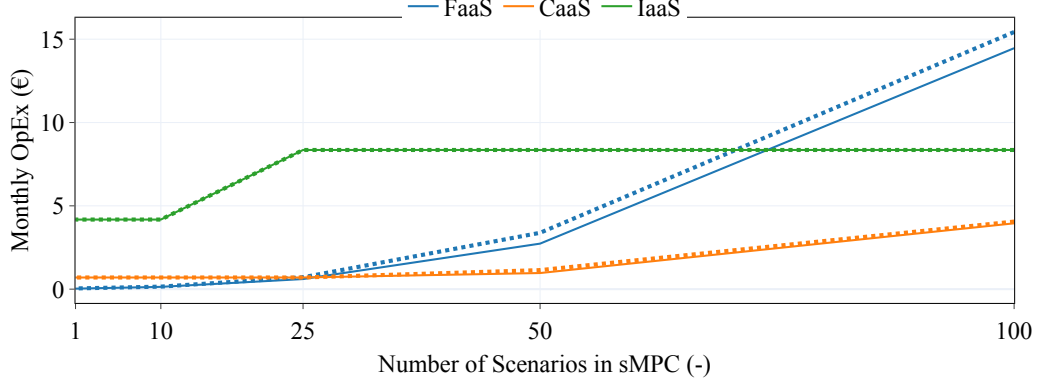


Figure 13: Sensitivity analysis of sMPC OpEx to the increasing number of scenarios, showing the lowest sensitivity for CaaS. Solid line represents cost of average workload while dotted shows maximum workload case.

Additionally, while commercial solvers are more computationally efficient than open-source alternatives, they introduce significant OpEx costs—often amounting to tens of thousands of euros annually. By using an open-source solver within a cloud-based architecture, the EMS provider achieves cost-effective scaling. These savings can be passed on to customers through lower service fees or more favorable profit sharing agreements, improving the business viability of the solution.

3.8. Profitability of EMS providers

It is important to distinguish between the financial outcomes of the microgrid operator and the EMS provider. In typical agreements, the EMS provider receives a percentage of the microgrid cost savings or additional revenue, commonly ranging between 10% and 20%. This revenue must cover the EMS provider’s OpEx for cloud and infrastructure services. Figure 14 illustrates the revenue-sharing mechanism, emphasizing how profitability scales with the number of aggregated microgrids.

As detailed in Section 3.7.2, operational costs remain low relative to the revenue generated for microgrid operators. However, the profitability of the EMS provider depends on cost-efficient scaling. Under a profit-sharing model, minimizing OpEx is essential to sustain margins while managing a growing portfolio of microgrids.

To quantify the impact of OpEx on EMS provider profitability, we compare our hybrid approach with a scenario where the EMS is fully deployed via

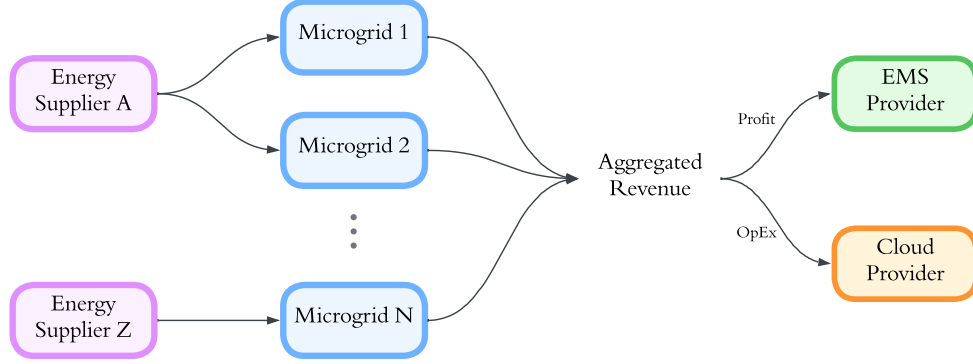


Figure 14: Profit-sharing mechanism between the microgrid operator and the EMS provider. The microgrid receives an energy bill from the supplier and retains a portion of the savings, while the EMS provider’s earnings depend on the number of managed microgrids, control performance, and operational costs.

IaaS. Based on the profit distribution in Table 5, the EMS provider generates 185.60 €/month per microgrid, of which 6% (11.37 €) is allocated to IaaS cloud provisioning. By adopting the proposed hybrid approach, OpEx drops tenfold to 1.52 € (0.8% of revenue), dramatically improving cost efficiency.

4. Conclusions

This paper presented a comprehensive energy management framework for microgrids, integrating advanced control methodologies with practical deployment considerations. The proposed EMS framework is designed with a layered control structure, where each operates on different timescales. The energy management layer is discussed in detail, as it plays a crucial role in optimizing power flows considering multiple revenue streams. This layer employs a scenario-based model predictive control approach to address the inherent uncertainties in microgrid systems. The MPC framework incorporates adjustable risk aversion, allowing the system to balance between safer conservative and riskier profit-seeking control strategies based on the user’s preferences. To enhance computational efficiency, a novel approximation method is introduced. The method suggests using stochastic MPC for generating parameters for a faster deterministic MPC. This approach ensures that the control actions are both robust to uncertainty and computationally

feasible, enabling real-time decision-making and adaptability to dynamic environments.

Beside elaborated model predictive control, the proposed framework incorporates a software architecture designed for practical implementation, forming a fully deployable EMS solution that can be offered as a service for microgrid operators. The EMS has been successfully deployed and operated in multiple microgrid facilities with different compositions and operational requirements, showcasing its versatility and scalability. In the examined case study, the EMS took advantage of energy market opportunities and generated profit exceeding 3000 €/month, which translates to a normalized yield of storage rated power to 159 €/kW/year. These results underscore the economic viability of the proposed EMS, reinforcing its potential to make microgrid operations not only environmentally sustainable but also financially attractive. In conclusion, the proposed EMS framework and software architecture provide a resilient and scalable solution for microgrid energy management, ensuring profitability and operational reliability in diverse applications.

Acknowledgement

The authors gratefully acknowledge the contribution of the Scientific Grant Agency of the Slovak Republic under the grant 1/0490/23 and the Slovak Research and Development Agency under the project APVV-20-0261.

Appendix A. Derivation of CVaR formulation

Let $J(s)$ denote the total cost associated with scenario $s \in \mathcal{S}$. The Value-at-Risk (VaR) at confidence level $\alpha \in (0, 1)$ is defined as

$$\text{VaR}_\alpha \triangleq \min\{y \in \mathbb{R} : \mathcal{P}(J(s) \leq y) \geq \alpha\}. \quad (\text{A.1})$$

While VaR captures the worst loss not exceeded with probability α , it is non-convex. The Conditional Value-at-Risk instead reflects the expected cost in the worst $(1 - \alpha)$ -tail:

$$\text{CVaR}_\alpha \triangleq \mathbb{E}[J(s) \mid J(s) \geq \text{VaR}_\alpha]. \quad (\text{A.2})$$

A convex reformulation is available through linear programming [37]:

$$\min_{y,z} \quad y + \frac{1}{1-\alpha} \sum_{s \in \mathcal{S}} \pi(s) z(s) \quad (\text{A.3a})$$

$$\text{s.t.} \quad z(s) \geq J(s) - y, \quad \forall s \in \mathcal{S}, \quad (\text{A.3b})$$

$$z(s) \geq 0, \quad \forall s \in \mathcal{S}, \quad (\text{A.3c})$$

where y approximates VaR, $z(s)$ are auxiliary variables, and $\pi(s)$ are scenario probabilities with $\sum_s \pi(s) = 1$. In this work, we assume equal scenario probabilities $\pi(s) = 1/N_S$.

References

- [1] M. A. Alarcón, R. G. Alarcón, A. H. González, A. Ferramosca, Economic model predictive control for energy management of a microgrid connected to the main electrical grid, *Journal of Process Control* 117 (2022) 40–51.
- [2] R. Silva, C. Gouveia, L. Carvalho, J. Pereira, Improved battery storage systems modeling for predictive energy management applications, in: 2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), IEEE, 2022, pp. 1–5.
- [3] T. Siqin, S. He, B. Hu, X. Fan, Shared energy storage-multi-microgrid operation strategy based on multi-stage robust optimization, *Journal of Energy Storage* 97 (2024) 112785.
- [4] A. Alizadeh, I. Kamwa, A. Moeini, S. M. Mohseni-Bonab, Energy management in microgrids using transactive energy control concept under high penetration of renewables; a survey and case study, *Renewable and Sustainable Energy Reviews* 176 (2023) 113161. doi:<https://doi.org/10.1016/j.rser.2023.113161>.
URL <https://www.sciencedirect.com/science/article/pii/S1364032123000175>
- [5] W. Yang, H. Fang, D. Xu, B. Jiang, P. Shi, A stochastic model predictive control-based energy management approach for microgrids with electric vehicles, *IEEE Transactions on Transportation Electrification* 11 (1) (2025) 3137–3145. doi:10.1109/TTE.2024.3435426.

- [6] V. Hamdipoor, H. N. Nguyen, B. Mekhaldi, J. Parra, J. Badosa, F. C. Obaldia, Experimental validation of scenario-based stochastic model predictive control of nanogrids, *Control Engineering Practice* 157 (2025) 106249.
- [7] Z. Cheng, D. Jia, Z. Li, S. Xu, C. Zhi, L. Wu, Multi-time scale energy management of microgrid considering the uncertainties in both supply and demand, *Energy Reports* 8 (2022) 10372–10384.
- [8] M. Mollayousefi Zadeh, P. MohammadAli Rezayi, S. Ghafouri, M. Alizadeh, G. Gharehpetian, Iot-based stochastic ems using multi-agent system for coordination of grid-connected multi-microgrids, *International Journal of Electrical Power & Energy Systems* 151 (2023) 109191. doi:10.1016/j.ijepes.2023.109191.
URL <https://www.sciencedirect.com/science/article/pii/S014206152300248X>
- [9] R. Smets, K. Bruninx, J. Bottieau, J.-F. Toubéau, E. Delarue, Strategic implicit balancing with energy storage systems via stochastic model predictive control, *IEEE Transactions on Energy Markets, Policy and Regulation* 1 (4) (2023) 373–385.
- [10] L. Herre, Risk-averse aggregator of controllable loads as virtual battery providing multiple services, in: 2020 17th International Conference on the European Energy Market (EEM), IEEE, 2020, pp. 1–6.
- [11] L. Herre, Impact of imbalance settlement system design on risk-averse energy storage, in: 2020 17th International Conference on the European Energy Market (EEM), IEEE, 2020, pp. 1–6.
- [12] Y. Wu, B. Xu, J. Anderson, Energy storage arbitrage under price uncertainty: Market risks and opportunities, *arXiv preprint arXiv:2501.08472* (2025).
- [13] M. Fochesato, C. Peter, L. Morandi, J. Lygeros, Peak shaving with hydrogen energy storage: From stochastic control to experiments on a 4 mwh facility, *Applied Energy* 376 (2024) 123965.
- [14] S. Rafayal, A. Alnaggar, M. Cevik, Optimizing electricity peak shaving through stochastic programming and probabilistic time series forecasting, *Journal of Building Engineering* 88 (2024) 109163.

- [15] S. Matrone, A. Pozzi, E. Ogliari, S. Leva, Deep learning-based predictive control for optimal battery management in microgrids, *IEEE Access* (2024).
- [16] V. Trovato, B. Kantharaj, Energy storage behind-the-meter with renewable generators: Techno-economic value of optimal imbalance management, *International Journal of Electrical Power & Energy Systems* 118 (2020) 105813.
- [17] H. Fan, Q. Yuan, H. Cheng, Multi-objective stochastic optimal operation of a grid-connected microgrid considering an energy storage system, *Applied Sciences* 8 (12) (2018) 2560.
- [18] A. C. Real, G. P. Luz, J. Sousa, M. Brito, S. Vieira, Optimization of a photovoltaic-battery system using deep reinforcement learning and load forecasting, *Energy and AI* 16 (2024) 100347.
- [19] G. P. Georgiadis, C. N. Dimitriadis, N. Passalis, M. C. Georgiadis, A hybrid ml-milp framework for the optimal integration of photovoltaic and battery systems in manufacturing industries, *Computers & Chemical Engineering* 203 (2025) 109356. doi:<https://doi.org/10.1016/j.compchemeng.2025.109356>.
URL <https://www.sciencedirect.com/science/article/pii/S0098135425003588>
- [20] M. Habib, E. Bollin, Q. Wang, Edge-based solution for battery energy management system: Investigating the integration capability into the building automation system, *Journal of Energy Storage* 72 (2023) 108479. doi:<https://doi.org/10.1016/j.est.2023.108479>.
URL <https://www.sciencedirect.com/science/article/pii/S2352152X23018765>
- [21] H. M. Yassim, M. N. Abdullah, C. K. Gan, A. Ahmed, A review of hierarchical energy management system in networked microgrids for optimal inter-microgrid power exchange, *Electric Power Systems Research* 231 (2024) 110329. doi:<https://doi.org/10.1016/j.epsr.2024.110329>.
URL <https://www.sciencedirect.com/science/article/pii/S0378779624002177>

- [22] R. Bustos, L. G. Marín, A. Navas-Fonseca, L. Reyes-Chamorro, D. Sáez, Hierarchical energy management system for multi-microgrid coordination with demand-side management, *Applied Energy* 342 (2023) 121145. doi:<https://doi.org/10.1016/j.apenergy.2023.121145>.
URL <https://www.sciencedirect.com/science/article/pii/S0306261923005093>
- [23] F. Karimi Pour, P. Segovia, E. Duviella, V. Puig, A two-layer control architecture for operational management and hydroelectricity production maximization in inland waterways using model predictive control, *Control Engineering Practice* 124 (2022) 105172. doi:<https://doi.org/10.1016/j.conengprac.2022.105172>.
URL <https://www.sciencedirect.com/science/article/pii/S0967066122000697>
- [24] J. Wu, S. Li, A. Fu, M. Cvetković, P. Palensky, J. C. Vasquez, J. M. Guerrero, Hierarchical online energy management for residential microgrids with hybrid hydrogen-electricity storage system, *Applied Energy* 363 (2024) 123020. doi:<https://doi.org/10.1016/j.apenergy.2024.123020>.
URL <https://www.sciencedirect.com/science/article/pii/S0306261924004033>
- [25] X. Lu, W. Gang, S. Cai, Z. Tu, Energy management strategy for a novel multi-stack integrated hydrogen energy storage system based on hybrid rules and optimization, *Applied Energy* 381 (2025) 125189. doi:<https://doi.org/10.1016/j.apenergy.2024.125189>.
URL <https://www.sciencedirect.com/science/article/pii/S030626192402573X>
- [26] L. Wu, X. Yin, L. Pan, J. Liu, Smart energy management: Process structure-based hybrid neural networks for optimal scheduling and economic predictive control in integrated systems, *Applied Energy* 380 (2025) 124965. doi:<https://doi.org/10.1016/j.apenergy.2024.124965>.
URL <https://www.sciencedirect.com/science/article/pii/S0306261924023481>
- [27] P. Ge, D. Tang, Y. Yuan, J. M. Guerrero, E. Zio, A hierarchical multi-objective co-optimization framework for sizing and energy management of coupled hydrogen-electricity energy

- storage systems at ports, *Applied Energy* 384 (2025) 125451.
doi:<https://doi.org/10.1016/j.apenergy.2025.125451>.
URL <https://www.sciencedirect.com/science/article/pii/S0306261925001813>
- [28] T. R. Nudell, M. Brignone, M. Robba, A. Bonfiglio, G. Ferro, F. Delfino, A. M. Annaswamy, Distributed control for polygeneration microgrids: A dynamic market mechanism approach, *Control Engineering Practice* 121 (2022) 105052. doi:<https://doi.org/10.1016/j.conengprac.2021.105052>.
URL <https://www.sciencedirect.com/science/article/pii/S0967066121002999>
- [29] F. Moazzen, M. Hossain, A two-layer strategy for sustainable energy management of microgrid clusters with embedded energy storage system and demand-side flexibility provision, *Applied Energy* 377 (2025) 124659. doi:<https://doi.org/10.1016/j.apenergy.2024.124659>.
URL <https://www.sciencedirect.com/science/article/pii/S0306261924020427>
- [30] Z. Wang, S. Yang, X. Xiang, A. Vasiljević, N. Mišković, Đula Nađ, Cloud-based mission control of usv fleet: Architecture, implementation and experiments, *Control Engineering Practice* 106 (2021) 104657. doi:<https://doi.org/10.1016/j.conengprac.2020.104657>.
URL <https://www.sciencedirect.com/science/article/pii/S0967066120302276>
- [31] M. Amoretti, R. Pecori, Y. Protskaya, L. Veltri, F. Zanichelli, A scalable and secure publish/subscribe-based framework for industrial iot, *IEEE Transactions on Industrial Informatics* 17 (6) (2021) 3815–3825. doi:[10.1109/TII.2020.3017227](https://doi.org/10.1109/TII.2020.3017227).
- [32] A. Banjac, H. Novak, M. Vašak, Implementation of model predictive indoor climate control for hierarchical building energy management, *Control Engineering Practice* 136 (2023) 105536. doi:<https://doi.org/10.1016/j.conengprac.2023.105536>.
URL <https://www.sciencedirect.com/science/article/pii/S0967066123001053>
- [33] S. A. G. K. Abadi, S. I. Habibi, T. Khalili, A. Bidram, A model pre-

- dictive control strategy for performance improvement of hybrid energy storage systems in dc microgrids, *IEEE Access* 10 (2022) 25400–25421.
- [34] D. Pozo, Linear battery models for power systems analysis, *Electric Power Systems Research* 212 (2022) 108565.
 - [35] Á. Tereza, R. Kohút, K. Fedorová, M. Kvasnica, Risk-aware stochastic energy management of microgrid with battery storage and renewables, *IFAC-PapersOnLine* 56 (2) (2023) 8445–8450.
 - [36] Q. Ni, S. Zhuang, H. Sheng, G. Kang, J. Xiao, An ensemble prediction intervals approach for short-term pv power forecasting, *Solar Energy* 155 (2017) 1072–1083.
 - [37] R. T. Rockafellar, S. Uryasev, et al., Optimization of conditional value-at-risk, *Journal of risk* 2 (2000) 21–42.
 - [38] Y. Hu, M. Armada, M. J. Sánchez, Potential utilization of battery energy storage systems (bess) in the major european electricity markets, *Applied Energy* 322 (2022) 119512.