Adaptive Demand-Side Management using Reinforcement Learning and Model Predictive Control under Electricity Market Uncertainty

Team Members: Zhe Li, Alexandra Janey, Amin Nassaji

Spring 2025

Executive Summary:

Industrial demand-side management (DSM) presents significant economic opportunities for energy-intensive industries participating actively in electricity markets. However, the integration of renewable energy sources into spot markets causes drastic fluctuations, introducing uncertainty to traditional load-shifting strategies. Conventional optimization techniques, such as stochastic programming and robust optimization, are often too computationally expensive for real-time applications, limiting their practical use in rapidly changing market environments.

In this project, the scheduling of a generic industrial process, represented as a flexible battery storage system, was modeled as a sequential decision-making problem and solved using reinforcement learning (RL). The RL agent was trained on day-ahead electricity price data from the German electricity spot market in 2017. To benchmark the RL agent's performance, a model predictive control (MPC) approach was also developed, using a neural network for short-term price forecasting.

It was shown that while a mixed-integer linear programming (MILP) approach achieves optimal costs with perfect price foresight, it lacked real-time adaptability. The MPC controller dynamically adjusted consumption based on forecasts but was sensitive to prediction errors and was computationally expensive. The RL agent, despite requiring extensive training, learned adaptive policies directly from price patterns and achieved near-optimal cost without relying on explicit future predictions. Overall, the project demonstrated the potential of RL as a flexible and scalable alternative to traditional optimization methods for real-time iundustrial DSM under market uncertainty.

1 Background

The energy sector is characterized by fluctuations and variability that challenge grid stability and sustainability. Industrial demand-side management (DSM) addresses these challenges by adjusting electricity consumption patterns of energy-intensive processes based on price fluctuations. This allows for these processes to participate more effectively in electricity markets, reduce the operational costs, and facilitate the integration of renewable energy sources. Electricity markets like Germany's spot market exhibit high price volatility which creates both economic opportunities and operational uncertainties for industries that can flexibly shift their loads.

Schäfer et al.[1] explored not only the economic benefits but also the environmental benefits of DSM using a generalized process model to simulate load-shifting. Their study utilized mixed-integer linear programming (MILP) to schedule processes optimally based on historic and projected electricity price data. They identified significant potential in reducing costs by leveraging daily fluctuations in electricity prices. However, this deterministic approach was reliant on day-ahead market data which limits the model's responsiveness to real-time market conditions.

Germscheid et al.[2] enhanced this model by formulating DSM as a two-stage stochastic program. Their model accounted for uncertainties in both day-ahead and same-day market prices. While this two-stage approach significantly improved decision-making, it did not have the capability to dynamically adjust decisions as real-time market conditions evolved.

To address these limitations, this project modeled DSM as a sequential decision-making process using reinforcement learning (RL) which enables the model to make real-time decision adjustments based on new market information. To benchmark the performance of the RL model, a model predictive control (MPC) approach was also developed. By comparing RL and MPC across rolling forecasts and real-time price uncertainty, the potential advantages of data-driven decision-making over traditional and computationally expensive optimization-based scheduling was assessed.

2 Data Description

Publicly available day-ahead electricity price data from Agora Energiewende[3] was used for model formulation and testing. The dataset was from the German electricity spot market for the year 2017 and contains 8,760 hourly price points which correspond to each hour of the year. Each data point represents the market clearing price in euros per megawatt-hour (EUR/MWh) for each given hour. The dataset is one-dimensional as it consists only of time-series price data and does not have any missing entries or gaps across the entire year. The model was trained with hours 3000 - 5000 as it was one of the most stable periods in the data and hours 5880 - 6048 were used for testing. By using this data to train and test the model, it was assumed that Germany's 2017 market behavior is representative of typical operating conditions.

3 Methods

3.1 Formulation of Demand-side Management Environment

DSM problem was modeled as a Markov decision process (MDP) and solved via RL. It is assumed that a process plant, which is abstracted as a battery with limited storage capacity and constant power demand, should decide the amount of power purchase in a fluctuating electricity spot market, taking advantage of flexible productions. The decisions were made on an hourly basis, since this is the frequency with which prices are updated.

3.1.1 State Variables

The state variables are designed to thoroughly capture the current situation of the environment. In this model, four state variables were tracked, forming a 4-dimensional vector space:

$$S_t = [s_t, p_{t-1}, c_t, \Delta c_t] \tag{1}$$

where t is the index for time in hours. $s_t \in [s^{\min}, s^{\max}]$ is the storage level and p_{t-1} is the amount of electricity purchased in the last time step. The spot market price c_t is obtained from historical data, and the rate of local price change is measured by $\Delta c_t = c_t - c_{t-1}$. These variables are provided to the RL agent to inform decision-making.

3.1.2 Action Variable

The only action variable in the model is the amount of electricity to purchase:

$$A_t = [p_t] \tag{2}$$

Since most chemical plants are designed for steady operation, the model assumes a nominal power uptake of $\bar{p} = 1.0$ as the constant power demand. The minimal and maximal flexible production scales are defined as $\theta_{\min}$ and $\theta_{\max}$, respectively. The power uptake p_t is constrained within the range $[\bar{p}\theta_{\min}, \bar{p}\theta_{\max}]$ due to the flexibility capacity of the process model. Furthermore, a nonlinear efficiency loss was considered at off-design operation. Following the approach from Schäfer et al.[1], the effective production rate p_t^- is estimated from a cubic function of the power uptake p_t:

$$p_t^- = \left[1 - \zeta \cdot \left(\frac{p_t - \bar{p}}{p_{\min} - \bar{p}}\right)^2\right] p_t \tag{3}$$

where the process parameter ζ characterizes the relative loss of production efficiency at minimal part-load $\theta_{\min}$ compared to the nominal power uptake $\bar{p}$.

3.1.3 Stage Cost

The stage cost consists of the cost of electricity purchases and penalties for breaking the constraints on ramping limit and storage capacity:

$$C_t = c_t \cdot p_t + Penalty_{ramp} + Penalty_{storage} \tag{4}$$

where α is the adjustable weighting coefficient. The ramping rate, defined as $\Delta_t = p_t - p_{t-1}$, measures the change in power uptake between successive time steps and is constrained within a safe range $[-\Delta_t^{\max}, \Delta_t^{\max}]$. The penalty for violating ramping limit constraints is expressed as:

$$Penalty_{ramp} = \alpha \cdot \max(|\Delta_t| - \Delta_t^{\max}, 0) \tag{5}$$

Similarly, an understock penalty is defined when the storage level is insufficient to satisfy the constant demand:

$$Penalty_{storage} = \alpha \cdot \max(s^{\min} - (s_t + p_t^- - \bar{p}), 0) \tag{6}$$

The objective of the sequential decision-making task is to minimize the discounted sum of future costs:

$$J = \sum_{t=1}^{N} \gamma^t c_t, \quad \gamma \in [0, 1] \tag{7}$$

where γ is the discount factor which assigns a higher weight to immediate costs than to future costs in the series.

3.1.4 Transition Function

At the end of each time step, the storage level is updated based on the effective production rate and demand satisfaction:

$$s_t \leftarrow s_t + p_t^- - \bar{p} \tag{8}$$

and the electricity price is updated with historical data.

3.2 Reinforcement Learning Algorithm: Proximal Policy Optimization

To handle continuous state and action spaces, the Proximal Policy Optimization (PPO) algorithm was applied to solve the DSM model. PPO improves traditional policy gradient methods by stabilizing learning through a clipped objective function, which ensures that the policy updates do not deviate significantly from the current policy. PPO utilizes neural network approximators for both the policy (actor) and the value function (critic). The actor network outputs a stochastic policy $\pi(a|s;\theta)$, while the critic network estimates the state

value function $V(s;\phi)$. The policy update is performed using the clipped surrogate objective:

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t \right) \right] \tag{9}$$

where $r_t(\theta) = \frac{\pi\theta(a_t|s_t)}{\pi\theta_{\text{old}}(a_t|s_t)}$ is the ratio of the new and old policy probabilities, $\hat{A}_t$ is the advantage estimate, and ϵ is the clipping threshold. Further details about the PPO algorithm can be referred to in the original paper[4].

The off-the-shelf PPO algorithm from the Stable-Baselines3 library[5], a commonly used RL library in PyTorch, was implemented. The hyperparameters are summarized in Table 1.

Hyperparameter	Description	Value
policy	The policy model to use	'MlpPolicy'
learning_rate	The learning rate	3e-4
gamma	The discount factor	0.995

Table 1: Important hyperparameters of PPO.

3.3 Model Predictive Control Approach

To benchmark the performance of the RL agent, a MPC baseline was developed. MPC optimized electricity consumption over a receding horizon using MILP, applying only the first control action at each time step before re-optimizing, guided by neural network-based price forecasts generated form historical data. The objective is to minimize total electricity cost while satisfying system constraints related to storage capacity, ramping limits, and production capacity.

The optimization problem is formulated as a MILP based on the work of Schäfer et al.[1], which models flexible electricity usage in industrial settings under dynamic pricing. The same constraint structure is adopted to ensure a consistent comparison with the RL agent. The problem is modeled using JuMP in Julia and solved using IPOPT[6]. This forms a hybrid control approach that combines machine learning (for forecasting) and optimization-based control (for decision-making).

3.3.1 Neural Network Forecasting Model

MPC relies on forecasts of future electricity prices, which are generated using a three-layer fully connected neural network. The model takes the previous 169 hours of prices as input and predicts the next 168 hours. Input data is normalized using Min-Max scaling, and training sequences are created using a sliding window approach. The network uses LeakyReLU activation, MSE loss, and the Adam optimizer.

4 Results

The behavior of the four different control strategies, Oracle (the original MILP), RL agent, MPC, and Steady baseline, is illustrated in Figure 1

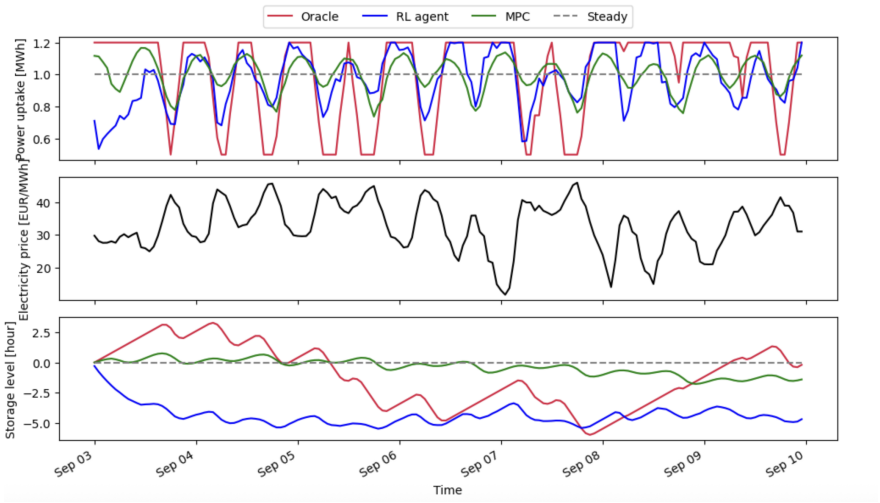

Figure 1: Results comparison of Oracle, RL, MPC, and Steady baseline models

The Oracle controller (red) uses complete knowledge of future electricity prices to compute an optimal open-loop schedule. It performs aggressive on/off switching to minimize costs by fully exploiting low-price periods. In contrast, the RL agent (blue), generates a smoother response. Although it does not have access to future prices, it learns to increase consumption during price dips and conserve energy when prices rise, based on past experience and environmental feedback. The MPC controller (green) operates using predicted prices and produces a more tempered power signal. The Steady baseline (dashed) maintains constant consumption at 1.0 MWh and serves as a non-optimized reference.

In terms of storage behavior, the RL agent diverges significantly due to the absence of explicit constraints on cumulative storage. While it maintains operational continuity, it tends to under-consume relative to the nominal demand. This behavior is expected in model-free learning unless explicit penalties or terminal storage goals are incorporated during training. Nonetheless, the RL agent achieves nearly optimal cost while exhibiting smooth and adaptive behavior.

4.1 Model Predictive Control Forecasting Performance

To evaluate the performance of the MPC controller, the accuracy of its price forecasts is examined. Figure 2 compares the predicted electricity prices (dashed orange line) from the neural network model to the actual prices (solid blue line) over a representative horizon.

The neural network captures the general trend and periodic behavior of the real market signal but exhibits phase shifts and amplitude mismatches. These discrepancies reflect the challenges of accurate time-series forecasting, particularly in highly variable and spiky domains like electricity pricing. Prediction accuracy of the MPC model is quantified by root

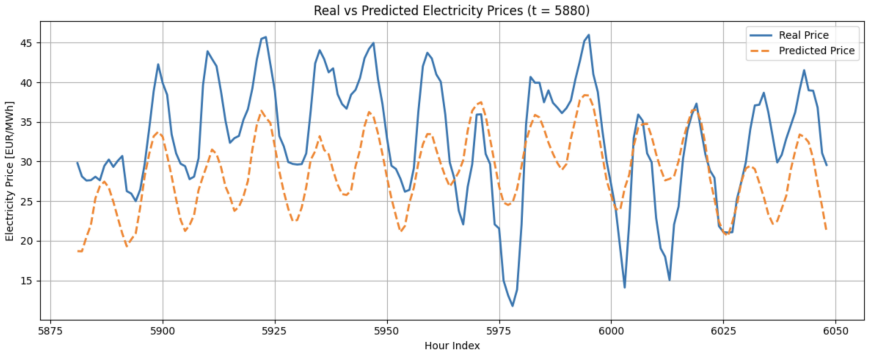

Figure 2: Price comparison of MPC predicted prices and real data prices

mean squared error of 7.2718 EUR/MWh and a mean absolute error of 6.4318 EUR/MWh. These values indicate a reasonable level of accuracy for the neural network price predictor, supporting the MPC's ability to make informed decisions despite imperfect forecasts.

4.2 Cost Comparison

The total cost incurred by each control method, including a penalty for final storage imbalance, is shown in Figure 3 below.

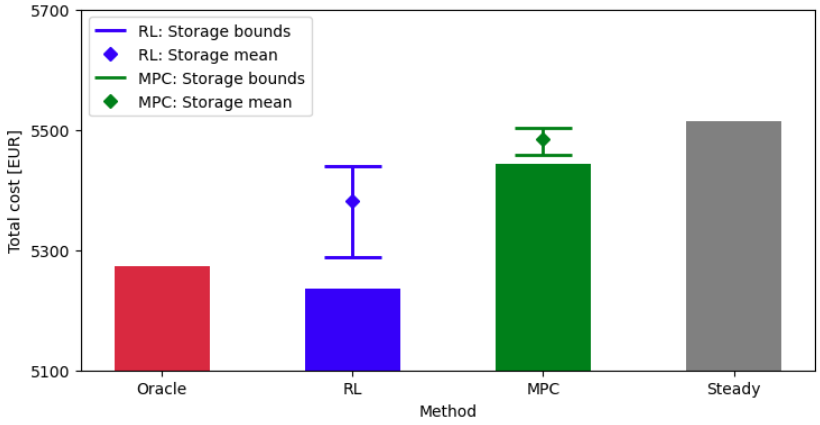

Figure 3: Cost comparison of Oracle, RL, Steady, and MPC approaches

The bar plot includes penalties for storage imbalance: if the final storage level deviates from zero, an equivalent energy cost is added. This cost is computed by multiplying the absolute storage imbalance by the minimum, average, and maximum prices over the time horizon which is visualized as error bars on the bar plot. The Oracle incurs the lowest cost by construction, as it uses perfect foresight. The RL agent achieves nearly optimal cost with a smoother control policy, despite its storage imbalance penalty. The MPC controller reduces cost from Steady baseline operation, but is still above the upper bound cost for the RL agent. The Steady baseline, which does not adapt to price fluctuations, incurs the highest cost.

5 Discussion

This project compared MILP, MPC, and RL approaches for DSM under dynamic electricity pricing, emphasizing the trade-offs between optimality, adaptability, and computational feasibility.

The MILP approach provided a globally optimal solution over the full time horizon by solving the problem with perfect future price information. However, perfect foresight is rarely available in practical settings. Therefore, despite MILP confirming the potential economic benefits of DSM strategies, its applications are limited to offline or day-ahead planning. This highlights the inherent limitations of deterministic modeling as it relies on full price forecasts and has high computational complexity which is not well suited for volatile energy markets.

The MPC controller utilized short-term price predictions to adjust decisions dynamically, offering a more practical alternative to MILP. However, MPC's effectiveness was found to be highly sensitive to the accuracy of the neural network forecasts. Forecasting errors led to suboptimal electricity consumption decisions and slight energy imbalances, which were not explicitly penalized during optimization. Furthermore, MPC solves the MILP at each time step, so it is extremely computationally expensive and not feasible for real-world application. This highlights MPC's reliance on accurate forecasting which poses substantial operational risks given that forecasting will always have uncertainties.

The RL agent demonstrated the ability to adaptively manage power consumption based solely on observed prices and internal states without requiring explicit price forecasts. Compared to the other optimization-based approaches, RL's model-free framework aligned with the project's goal to achieve scalable, flexible control in uncertain markets. However, RL training required a significant computational effort and the resulting policies did not strictly enforce cumulative energy balance constraints unless explicitly trained to do so. These findings highlight RL's potential as a robust solution under market uncertainties, but operational reliability must be improved through extensive training to enforce cumulative constraints.

Looking forward, it is recommended that when applying RL to DSM strategies, the reward function should be modified to incorporate explicit penalties for the final storage imbalance. This will ensure better management of cumulative energy. For the MPC models, more advanced time-series techniques could be further investigated to better improve the forecast models. For example, Long Short-Term Memory networks[7] or Transformer architectures[8] could better improve predictive accuracy. Finally, hybrid control frameworks could be another promising modeling approach as they could integrate RL's adaptability with MPC's constraint enforcement capabilities. This could create a model that not only adapts dynamically to real-time market changes, but also maintains operational safety and reliability.

In conclusion, the results indicate that while traditional optimization methods establish optimal performance benchmarks, adaptive strategies such as RL have more practical advantages. RL's capability to operate effectively without explicit forecasting prices offers promising opportunities for real-time DSM. Integration of hybrid methods could further improve operational reliability and cost-effectiveness. This would offer more robust and practical solutions for real-time energy management in fluctuating electricity markets.

References

[1] Schäfer, P.; Daun, T. M.; Mitsos, A. Do investments in flexibility enhance sustainability? A simulative study considering the German electricity sector. *AIChE Journal* **2020**, *66*, e17010.

[2] Germscheid, S. H.; Mitsos, A.; Dahmen, M. Demand response potential of industrial processes considering uncertain short-term electricity prices. *AIChE Journal* **2022**, *68*, e17828.

[3] Agora Energiewende Historic time series data (2017). `https://www.agora-energiewende.de/`, 2017.

[4] Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347* **2017**,

[5] Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; Dormann, N. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research* **2021**, *22*, 1–8.

[6] Wächter, A.; Biegler, L. T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming* **2006**, *106*, 25–57.

[7] Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735–1780.

[8] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. Advances in Neural Information Processing Systems. Red Hook, NY, 2017; pp 5998–6008.