Towards Sustainable Hydrogen Systems: Supply Chain Optimization with Model Predictive Control and Reinforcement Learning

Mahammad Valiyev1

[1]Chemical Engineering, University of Southern California, Los Angeles, California, United States

Abstract

Hydrogen supply chains are expected to play a central role in future low-carbon energy systems. However, the operation of such systems is inherently challenged by factors such as uncertainty in renewable availability, electricity prices, and hydrogen demand, as well as by practical engineering constraints associated with electrolyzers, energy storage, and grid interaction. Effective control strategies must therefore balance economic performance, operational feasibility, and sustainability objectives under time-varying and uncertain conditions, especially as hydrogen infrastructure scales beyond pilot deployments.

This paper investigates and compares four operational control approaches for a renewable-powered hydrogen supply chain: a rule-based controller (RBC), model predictive control (MPC), reinforcement learning without forecasts (RL-NF), and reinforcement learning with forecast-augmented observations (RL-F). All methods are evaluated within a unified, physically realistic framework that incorporates electrolyzer minimum load and ramping constraints, battery and hydrogen storage dynamics, grid import limits, renewable curtailment, and consistent economic assumptions. This unified formulation ensures methodological consistency, enabling a fair and rigorous comparison of predictive and adaptive control paradigms under identical system conditions and uncertainty realizations.

Simulation results show that MPC achieves the highest economic performance by effectively exploiting short-term forecasts to reduce grid dependence and optimize storage utilization, highlighting the value of predictive optimization when reliable forecasts are available. RL-NF demonstrates robust and competitive performance despite operating without future information, while RL-F does not consistently outperform its no-forecast counterpart, illustrating the challenges of learning from noisy forecasts in complex energy systems. These findings illustrate the relative strengths and limitations of predictive and learning-based control approaches for hydrogen supply chains under uncertainty and motivate further investigation into their roles in large-scale hydrogen system operation.

Introduction

Global decarbonization agendas increasingly emphasize the need for scalable, low-carbon energy carriers capable of integrating high shares of renewable generation. In this context, hydrogen has emerged as a promising option for long-duration energy storage, industrial decarbonization, and clean fuel supply for transportation and power systems (IEA, 2019; Staffell et al., 2019). When produced using renewable electricity, hydrogen can enable deep emissions reductions while enhancing system flexibility in energy systems with growing shares of variable wind and solar generation. However, realizing this potential at scale requires not only infrastructure investment, but also effective operational strategies capable of managing uncertainty, variability, and complex physical constraints.

The operation of hydrogen supply chains presents significant challenges due to uncertainty in renewable availability, electricity market prices, and hydrogen demand, as well as engineering constraints associated with electrolyzers, energy storage systems, and grid interaction. Electrolyzers exhibit minimum loading and ramping constraints, energy storage assets are subject to capacity and throughput limits, and grid imports are often constrained by economic and regulatory considerations. As hydrogen systems scale beyond pilot projects, suboptimal operational decisions can lead to excessive grid dependence, renewable curtailment, increased costs, or underutilized assets. Consequently, optimizing hydrogen supply chain operation under uncertainty has become a critical research problem at the intersection of energy systems engineering, optimization, and control.

A wide range of control and optimization approaches have been explored for energy systems and energy supply chains. Rule-based control strategies remain common in practice due to their simplicity, interpretability, and ease of implementation, particularly in early-stage or pilot-scale deployments (Oldewurtel et al., 2012; Behzadi et al., 2023). However, such approaches are inherently reactive and typically struggle to exploit future information or adapt efficiently to changing operating conditions as system complexity increases. Model Predictive Control (MPC) has therefore received significant attention in recent years for hydrogen-integrated energy systems, including renewable microgrids, power-to-hydrogen facilities, and hybrid storage systems, owing to its ability to explicitly incorporate forecasts, operational constraints, and economic objectives within a receding-horizon optimization framework (Parisio et al., 2014; Rawlings et al., 2017; Birka et al., 2016; Velarde et al., 2017). Recent studies demonstrate that MPC can substantially reduce grid reliance, renewable curtailment, and operating costs in hydrogen-coupled systems when short-term forecasts are sufficiently reliable, though at the expense of increased computational burden and sensitivity to forecast errors (Huang et al., 2022).

More recently, reinforcement learning (RL) has emerged as a data-driven alternative for controlling complex energy and hydrogen systems under uncertainty, particularly in settings where accurate system models or reliable forecasts are difficult to obtain (Glavic et al., 2017; Vázquez-Canteli and Nagy, 2019). Advances in deep RL and safe learning frameworks have enabled applications to battery dispatch, integrated energy systems, and hydrogen production scheduling (Ruelens et al., 2017; Yang et al., 2021; Liang et al., 2024; Ye et al., 2024). These studies suggest that RL can learn adaptive policies capable of handling nonlinear dynamics and stochastic disturbances, but also highlight persistent challenges related to constraint satisfaction, training stability, and interpretability. In addition, recent work indicates that augmenting RL policies with forecast information does not necessarily guarantee improved performance, particularly when forecasts are noisy or only weakly informative (Oh et al., 2020). As a result, it remains an open question under which conditions forecast-augmented RL approaches can outperform reactive RL policies or established predictive optimization methods such as MPC in hydrogen supply-chain operations.

This paper contributes to the ongoing discussion by presenting a unified comparative study of four operational control strategies for a renewable-powered hydrogen supply chain: a rule-based controller (RBC), model predictive control (MPC), reinforcement learning without forecasts (RL-NF), and reinforcement learning with forecast-augmented observations (RL-F). This study presents a unified and constraint-consistent comparison of rule-based, predictive optimization, and forecast-augmented reinforcement learning controllers within a single hydrogen supply-chain framework, evaluated under identical disturbances, constraints, and economic assumptions.The modeled system captures the core elements of emerging green hydrogen configurations, including variable renewable generation, battery storage, an electrolyzer with minimum load and ramping constraints, hydrogen storage with throughput limits, grid interaction, and uncertain demand. This unified formulation enables a rigorous apples-to-apples comparison of predictive and learning-based control paradigms.

The results demonstrate clear differences in how predictive and adaptive control paradigms manage uncertainty and system constraints. MPC achieves the highest economic performance by effectively exploiting short-term forecasts to reduce grid reliance and optimize storage utilization. Reinforcement learning without forecasts exhibits robust and competitive performance despite operating under partial observability, learning conservative policies that prioritize renewable utilization and high hydrogen utilization rates. In contrast, reinforcement learning with forecast-augmented observations does not consistently outperform its no-forecast counterpart, highlighting the challenges associated with learning from noisy forecast information in complex, constrained energy systems. These findings illustrate the relative strengths and limitations of predictive and learning-based control approaches for hydrogen supply chains under uncertainty and motivate further investigation into robust and hybrid control strategies as hydrogen infrastructure scales.

Methodology

System Description and Problem Formulation

Hydrogen Supply Chain Process Model

This study considers a stylized renewable-powered hydrogen supply chain designed to capture the key operational interactions between variable renewable generation, energy storage, hydrogen production, and market demand. The system represents a typical emerging green hydrogen configuration and consists of four main components, illustrated in **Figure 1.** Figure 1 summarizes the governing equations and system architecture of the renewable-powered hydrogen supply chain. It links solar generation, battery storage, grid interaction, electrolyzer operation, hydrogen storage, and sales through energy balance relationships and operational constraints that define the feasible control space.

Renewable electricity source
Solar photovoltaic generation provides a variable, low-cost electricity input to the system. Solar availability is time-dependent and exhibits diurnal patterns with stochastic variability, representing typical photovoltaic generation behavior.

Battery energy storage system (BESS)
A battery is used to temporarily store electricity by charging during periods of excess solar generation and discharging when renewable availability is insufficient. The battery is characterized by finite power and energy capacity limits, as well as charging and discharging efficiencies.

Electrolyzer
The electrolyzer converts electrical energy into hydrogen with a fixed conversion efficiency. Electricity supplied to the electrolyzer can originate from solar generation, battery discharge, or the electricity grid. The electrolyzer is subject to rated power limits, minimum loading constraints, and ramping constraints.

Hydrogen storage and demand
Produced hydrogen can either be sold immediately to meet exogenous demand or stored in a hydrogen storage tank with limited capacity. Stored hydrogen may later be withdrawn to satisfy demand shortfalls, subject to storage throughput constraints.

The system operates in discrete hourly time steps over a one-week simulation horizon and is exposed to time-varying solar availability, electricity prices, and hydrogen demand.

State Variables, Control Variables, and System Dynamics

State Variables

System dynamics are governed by two primary state variables:

- **Battery state of charge (SOC):**

$$\mathrm{SOC}_t \in [0, \mathrm{SOC}_{\max}],$$

representing the stored electrical energy in the battery at time t.

- **Hydrogen storage level:**

$$H_t \in [0, H_{\max}],$$

representing the amount of hydrogen stored in the tank.

These states evolve according to energy and mass balance equations driven by control actions and system efficiencies.

Control Variables

At each time step, the controller determines energy and hydrogen allocation decisions, including:

- $E_{sb,t}$: electricity charged from solar into the battery
- $E_{be,t}$: electricity discharged from the battery to the electrolyzer
- $E_{ge,t}$: electricity drawn from the grid to the electrolyzer
- $E_{se,t}$: electricity supplied directly from solar generation to the electrolyzer.
- $P_{e,t}$: total electrical power supplied to the electrolyzer
- $H_{es,t}$: hydrogen sold directly from production
- $H_{est,t}$: hydrogen sent to storage
- $H_{ss,t}$: hydrogen sold from storage

In the MPC formulation, these variables are optimized explicitly. In the reinforcement learning formulations, high-level agent actions are mapped to feasible values of these variables through a feasibility layer, ensuring consistent system behavior and constraint satisfaction across all methods.

System Dynamics

The battery state of charge evolves according to:

$$\mathrm{SOC}_{t+1} = \mathrm{SOC}_t + \eta_{\mathrm{ch}} E_{sb,t} - \frac{E_{be,t}}{\eta_{\mathrm{dch}}}.$$

Hydrogen storage dynamics are given by:

$$H_{t+1} = H_t + H_{est,t} - H_{ss,t}.$$

Hydrogen production from the electrolyzer is modeled as:

$$H_{\mathrm{prod},t} = \eta_{\mathrm{elec}} P_{e,t}.$$

Total hydrogen sales satisfy:

$$H_{\mathrm{sales},t} = H_{es,t} + H_{ss,t}.$$

Objective Function

The operational objective is to maximize net profit over the simulation horizon, equivalently minimizing total operating cost minus hydrogen sales revenue. At each time step, the instantaneous cost includes:

- Grid electricity cost, including a carbon adder
- Battery cycling cost
- Electrolyzer operating cost
- Hydrogen storage holding cost

Revenue is generated through hydrogen sales.

The per-step objective can be written as:

$$\min \sum_t [(\pi_t + c_{\mathrm{CO_2}}) E_{ge,t} + c_{\mathrm{bat}}(E_{sb,t} + E_{be,t}) + c_{\mathrm{elec}} P_{e,t} + c_{\mathrm{store}} H_t - p_H H_{\mathrm{sales},t}].$$

This objective is used directly in the MPC formulation and serves as the reward signal for the reinforcement learning controllers, enabling consistent economic comparison across methods.

Operational Constraints

System operation is subject to the following constraints:

- Electrolyzer power limits:

$$0 \le P_{e,t} \le P_{\max}$$

$$P_{e,t} \ge P_{min} \cdot z_t$$

- Battery power and energy limits:

$$0 \le E_{sb,t}, E_{be,t} \le P_{\text{bat,max}}, 0 \le \text{SOC}_t \le \text{SOC}_{\text{max}}.$$

- Hydrogen storage capacity:

$$0 \le H_t \le H_{\text{max}}.$$

- Hydrogen balance:

$$H_{\text{prod},t} = H_{es,t} + H_{est,t}.$$

- Energy balance:

$$P_{\text{e},t} = E_{se,t} + E_{be,t} + E_{ge,t}.$$

- Demand satisfaction:

$$H_{\text{sales},t} \le D_t.$$

- Energy balance constraint:

$$E_{\text{se},t} + E_{\text{sb},t} \le S_t.$$

These constraints are enforced explicitly in the MPC formulation and implicitly through a feasibility layer in the reinforcement learning controllers, ensuring physically realistic and safe operation under all control strategiesUnmet hydrogen demand is permitted and does not incur an explicit penalty beyond foregone sales revenue. This assumption is adopted to simplify the economic formulation and isolate the effects of operational decision-making under uncertainty.

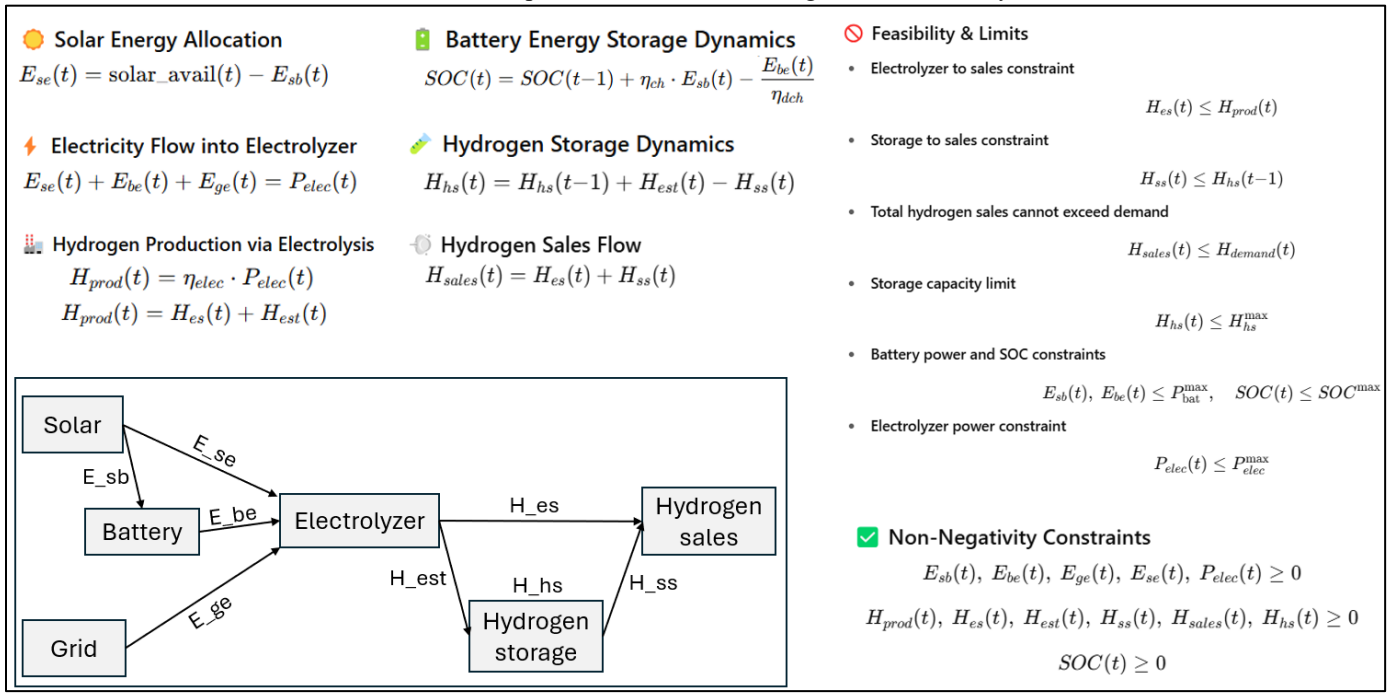


Figure 1. Process model, mass & energy balance, feasibility & limits

Disturbance and Forecast Models

The hydrogen supply chain is driven by three exogenous, time-varying disturbances evaluated at an hourly resolution $t = 0,1,\dots,T-1$. These disturbances are treated as uncontrollable inputs to the system and represent renewable availability, market conditions, and end-use demand. All controllers are evaluated under identical disturbance realizations to ensure a fair comparison.

Solar availability

Solar power availability S_t[kW] is modeled to capture the dominant diurnal cycle of photovoltaic (PV) generation, combined with slow multi-day variability and small stochastic fluctuations. Specifically,

$$S_t = \max\left(0,\, P_{\mathrm{PV}}^{\max}\, b(h_t)\,(c(d_t) + \varepsilon_t^{(S)})\right),$$

where $P_{\mathrm{PV}}^{\max} = 300$ kWis the installed PV capacity, $h_t = t \bmod 24$is the hour of day, and $d_t = \lfloor t/24 \rfloor$is the day index. The deterministic diurnal shape is given by a half-sine function

$$b(h) = \begin{cases} 0, & h < 6 \text{ or } h > 18, \\ \sin\left(\pi \dfrac{h-6}{12}\right), & 6 \le h \le 18, \end{cases}$$

while slow cloud-induced variability is modeled as

$$c(d) = 0.75 + 0.20\sin\left(\frac{2\pi d}{7}\right).$$

Small, zero-mean Gaussian noise $\varepsilon_t^{(S)} \sim \mathcal{N}(0, 0.03^2)$captures short-term irradiance fluctuations. The $\max(0,\cdot)$operator enforces non-negativity. The realized solar disturbance and representative short-horizon stochastic forecasts used by the predictive controllers are illustrated in Figure 2.

Electricity prices

The grid electricity price π_t[$/kWh] follows a time-of-use (TOU) tariff structure with occasional price spikes:

$$\pi_t = \pi_{\mathrm{TOU}}(h_t)\,\xi_t,$$

where the baseline TOU price is defined as

$$\pi_{\mathrm{TOU}}(h) = \begin{cases} 0.08, & 0 \le h < 6, \\ 0.12, & 6 \le h < 16, \\ 0.28, & 16 \le h < 21, \\ 0.15, & 21 \le h < 24, \end{cases}$$

and $\xi_t$is a stochastic spike multiplier:

$$\xi_t = \begin{cases} 2.5, & \text{with probability } 0.02, \\ 1, & \text{otherwise.} \end{cases}$$

This formulation reflects typical peak-hour pricing behavior and rare but impactful market price excursions. The realized electricity price disturbance and representative short-horizon stochastic forecasts are shown in Figure 3.

Hydrogen demand

Hydrogen demand D_t[kg/h] is modeled to exhibit daily consumption patterns, weekly modulation, and random variability:

$$D_t = \max\left(0,\, 3.5 + 1.5\sin\left(\frac{2\pi(h_t - 8)}{24}\right) + 0.3\sin\left(\frac{2\pi d_t}{7}\right) + \varepsilon_t^{(D)}\right),$$

where $\varepsilon_t^{(D)} \sim \mathcal{N}(0, 0.25^2)$. The demand magnitude is intentionally scaled to be consistent with a 300 kW electrolyzer, corresponding to a maximum production rate of approximately 6 kg/h given a conversion factor of 50 kWh/kg.

The realized hydrogen demand disturbance and representative short-horizon stochastic forecasts are shown in Figure 4.

Forecast availability and uncertainty

Disturbances are assumed to be known perfectly only at the current time step. **Short-horizon stochastic forecasts** are available to the MPC controller and the forecast-augmented reinforcement learning controller, while the no-forecast RL controller operates under partial observability.

For any disturbance signal $x_t \in \{S_t, \pi_t, D_t\}$, a forecast issued at time $t_0$for lead time $\delta = 1, \dots, H$is modeled as

$$\hat{x}_{t_0+\delta|t_0} = x_{t_0+\delta} + \eta_\delta, \eta_\delta \sim \mathcal{N}\,(0, \sigma_x^2(\delta)),$$

where $H = 12$hours denotes the prediction horizon.

The forecast uncertainty grows with lead time according to

$$\sigma_x(\delta) = \sigma_{x,0}\sqrt{1 + \alpha\delta}.$$

The base uncertainty levels are signal-specific:

$$\sigma_{S,0} = 10 \text{ kW},\ \sigma_{\pi,0} = 0.02\frac{\$}{kWh}, \sigma_{D,0} = 0.5 \text{ kg/h},$$

and $\alpha = 0.8$controls the rate of uncertainty growth with forecast horizon.

The forecast model is intentionally stylized and designed for controlled comparison rather than operational forecasting realism.

All forecasts are generated using identical disturbance models and random seeds across control strategies, ensuring that performance differences arise solely from control logic rather than informational asymmetry.

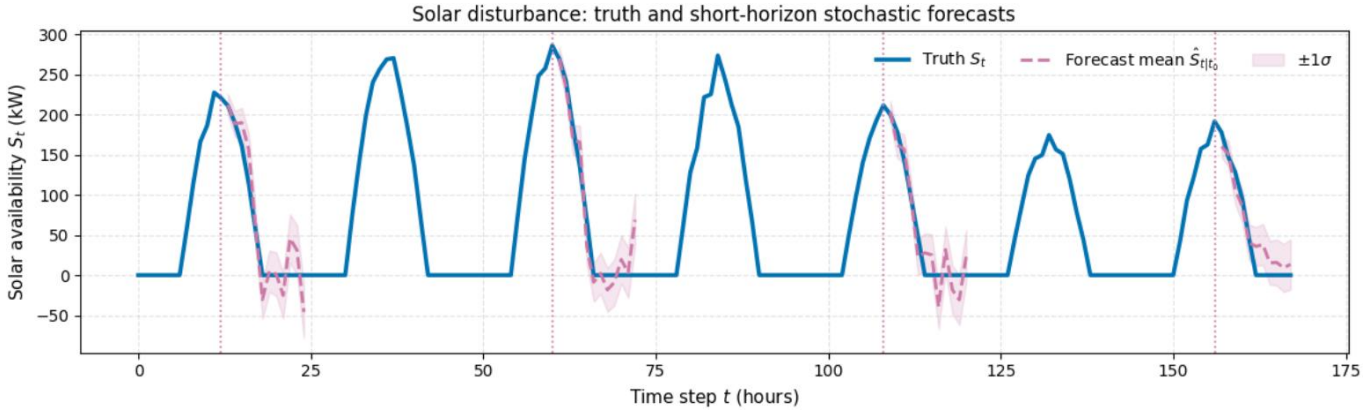


Figure 3. Solar disturbance: truth and short-horizon stochastic forecasts

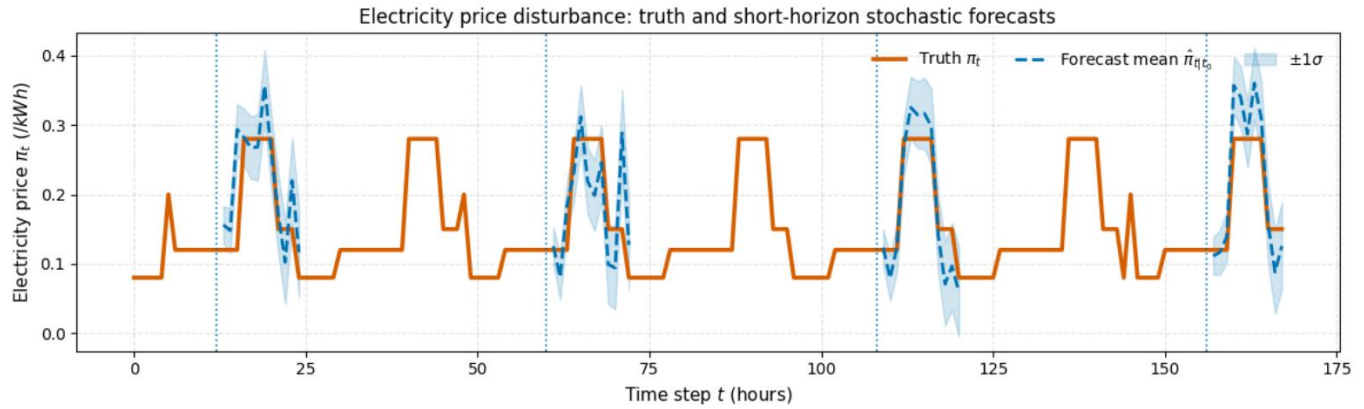


Figure 2. Electricity price disturbance: truth and short-horizon stochastic forecasts

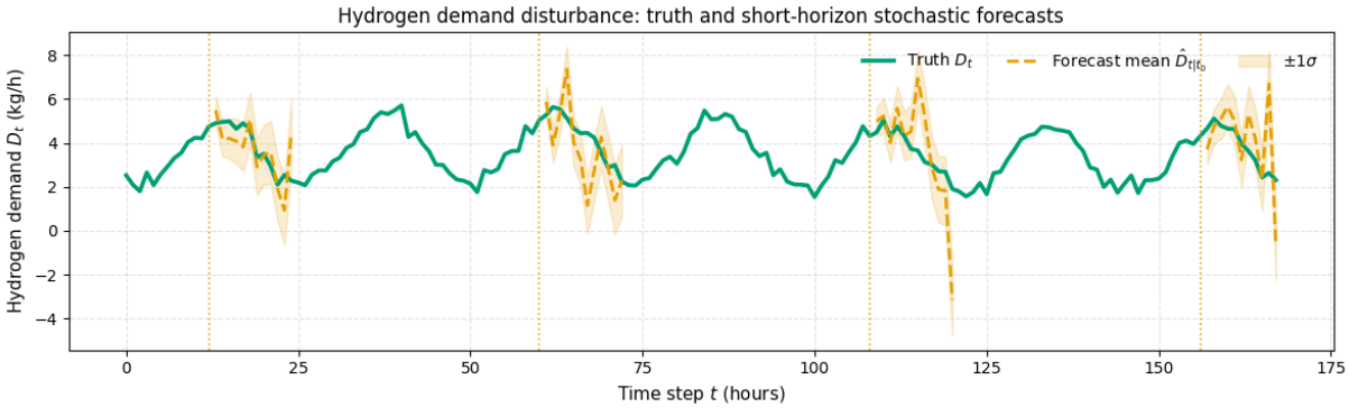


Figure 4. Hydrogen demand disturbance: truth and short-horizon stochastic forecasts

Control Strategies

This section presents the four control strategies evaluated in this study. All controllers operate on the same system model, constraints, cost structure, and disturbance realizations described in Section 2. Differences arise solely from how control decisions are generated, particularly in the use of forecasts, optimization, and learning. This design ensures a fair, apples-to-apples comparison.

Rule-Based Control (RBC)

Rule-based control (RBC) represents a heuristic, non-optimizing baseline commonly used in practice due to its simplicity and transparency. The controller operates myopically, relying only on current system states and disturbances, without forecasts or explicit optimization.

Logic Description

At each time step t, the RBC policy follows a fixed priority structure:

1. Electrolyzer dispatch

- Available solar energy is first allocated to the electrolyzer, up to its rated capacity $P_{\max}$.
- Excess solar, if available and if the battery is not near full charge, is used to charge the battery.

2. Battery usage

- If solar power is insufficient to reach $P_{\max}$and the battery state of charge satisfies

$$\mathrm{SOC}_t > 0.2\ \mathrm{SOC}_{\max},$$

the battery is discharged to supply the electrolyzer.

- Battery charging and discharging are constrained by power and energy limits.

3. Grid electricity

- Any remaining electrolyzer power requirement is met by drawing electricity from the grid, independent of electricity price.

4. Hydrogen allocation

- Hydrogen produced at time t,

$$H_{\mathrm{prod},t} = \eta_{\mathrm{elec}} P_{e,t},$$

is first used to meet current demand D_t.

- If production exceeds demand, surplus hydrogen is stored, subject to storage capacity.

- If production is insufficient, hydrogen is withdrawn from storage to meet demand when available.

Characteristics

- No forecasting: future solar, price, and demand information is ignored.
- No optimization: decisions are rule-driven rather than cost-optimal.
- Reactive behavior: actions respond only to current conditions.

Purpose

RBC serves as a benchmark baseline, providing a lower bound on performance against which more advanced control strategies can be evaluated.

Model Predictive Control (MPC)

Model Predictive Control (MPC) formulates system operation as a rolling-horizon optimization problem that explicitly accounts for system dynamics, constraints, and forecasts of future disturbances.

Rolling-Horizon Formulation

At each time step t, MPC solves a finite-horizon optimization problem over a prediction horizon H, using forecasts of solar availability $\hat{S}_{t+k}$, electricity prices $\hat{\pi}_{t+k}$, and hydrogen demand $\widehat{D}_{t+k}$for $k = 0, \dots, H-1$.

The optimization variables include:

$$\{E_{sb,t+k}, E_{be,t+k}, E_{ge,t+k}, P_{e,t+k}, H_{es,t+k}, H_{est,t+k}, H_{ss,t+k}\}.$$

State dynamics follow:

$$\mathrm{SOC}_{t+k+1} = \mathrm{SOC}_{t+k} + \eta_{\mathrm{ch}} E_{sb,t+k} - \frac{E_{be,t+k}}{\eta_{\mathrm{dch}}},$$
$$H_{t+k+1} = H_{t+k} + H_{est,t+k} - H_{ss,t+k}.$$

Objective Function

The MPC minimizes the cumulative economic cost over the horizon:

$$\min \sum_{k=0}^{H-1} [(\hat{\pi}_{t+k} + c_{\mathrm{CO_2}}) E_{ge,t+k} + c_{\mathrm{bat}}(E_{sb,t+k} + E_{be,t+k}) + c_{\mathrm{elec}} P_{e,t+k} + c_{\mathrm{store}} H_{t+k} - p_H H_{\mathrm{sales},t+k}].$$

This objective directly matches the formulation in Section 2.4, ensuring consistency with RL rewards.

Constraints

All physical and operational constraintsare enforced explicitly, including power limits, storage bounds, hydrogen balance, and demand satisfaction.

Receding-Horizon Execution

After solving the optimization problem, only the first control action is implemented. The horizon then shifts forward by one step, forecasts are updated, and the problem is re-solved at the next time step.

Characteristics:

- Explicit foresight via multi-step forecasts
- Optimization-based planning across time
- Guaranteed constraint satisfaction (subject to solver feasibility)

Reinforcement Learning (RL) Framework

Reinforcement learning controllers learn a control policy through interaction with the environment, without solving an explicit optimization problem at runtime.

Environment Definition:

The hydrogen supply chain is formulated as a Markov decision process (MDP), where:

- State: system states and exogenous information available to the agent
- Action: high-level continuous control vector
- Transition: governed by the physical dynamics in Section 2.3
- Reward: per-step economic profit

A feasibility layer maps agent actions to physically admissible control variables, ensuring consistency with system constraints across all RL variants.

State, Action, and Reward

- Action

$$a_t = [u_{pe,t}, u_{bat,t}, u_{store,t}] \in [0,1]^3,$$

controlling electrolyzer utilization, battery charging/discharging preference, and hydrogen storage preference.

- Reward
  The reward at each time step is defined as:

$$r_t = p_H H_{\text{sales},t} - [(\pi_t + c_{\text{CO}_2})E_{ge,t} + c_{\text{bat}}(E_{sb,t} + E_{be,t}) + c_{\text{elec}}P_{e,t} + c_{\text{store}}H_t],$$

identical to the MPC objective on a per-step basis.

Training Setup

All RL agents are trained using Proximal Policy Optimization (PPO) with continuous action spaces, identical hyperparameters, and identical disturbance realizations to ensure fair comparison.

Reinforcement Learning without Forecasts

In the no-forecast RL variant, the agent observes only current information, consisting of:

$$[\text{SOC}_t, H_t, S_t, \pi_t, D_t],$$

normalized to fixed ranges.

Partial Observability

Because future disturbances are not included in the observation, the agent operates under partial observability. Anticipating future price spikes, solar availability, or demand changes is not possible.

Reactive Learning

As a result, learned policies tend to be reactive, responding to immediate rewards rather than planning intertemporally. This can lead to conservative storage usage and reduced production scale, rather than reduced utilization.

Reinforcement Learning with Forecasts

The forecast-augmented RL variant extends the observation to include short-horizon noisy forecasts:

$$[\hat{S}_t, \hat{\pi}_t, \hat{D}_t, \dots, \hat{S}_{t+H_f-1}, \hat{\pi}_{t+H_f-1}, \hat{D}_{t+H_f-1}],$$

generated using the same stochastic forecast model employed by MPC.

Forecast-Augmented Observation

By augmenting the state with predicted future disturbances, the agent gains limited foresight, enabling it to learn anticipatory behaviors such as pre-charging the battery or deferring hydrogen sales.

Limited Foresight vs. MPC

Despite access to forecasts, forecast-augmented RL differs fundamentally from MPC:

- RL does not solve a multi-step optimization problem.
- Constraint coupling across time is implicit, not explicit.
- Planning emerges indirectly through policy learning rather than explicit reasoning.

Consequently, while forecast-augmented RL improves performance relative to reactive RL, it remains less effective than MPC in exploiting intertemporal trade-offs under uncertainty.

Feasibility Layer for Reinforcement Learning

In the reinforcement learning formulations, agent actions do not directly specify physical power or flow variables. Instead, actions are interpreted as normalized preferences and mapped deterministically to admissible control variables through a feasibility layer. This layer enforces all physical and operational constraints by (i) clipping control targets to allowable ranges, (ii) allocating available solar, battery, and grid power sequentially to satisfy electrolyzer demand, and (iii) ensuring energy and mass balance at each time step. As a result, all RL controllers operate on the same feasible domain as MPC, guaranteeing physically realistic operation and enabling a fair comparison across methods.

Figure 5 summarizes the structural differences among the four control strategies in terms of forecast usage, optimization, learning capability, and constraint handling. RBC relies on hard-coded rules without forecasts or learning, MPC leverages forecasts with explicit optimization, and the RL approaches rely on learning, with the forecast-augmented variant incorporating short-horizon predictions. Constraint enforcement differs accordingly, with MPC using explicit constraints and RL methods employing a feasibility layer.

Controller	Forecasts	Optimization	Learning	Constraint Handling
RBC	❌	❌	❌	Hard-coded
MPC	✅	✅	❌	Explicit
RL (no-forecast)	❌	❌	✅	Feasibility layer
RL (forecast)	✅	❌	✅	Feasibility layer

Figure 5. Summary of control strategies and information structure.

Results

Rule-Based Control (RBC): Baseline Performance

The rule-based controller (RBC) provides a reactive benchmark without forecasting or optimization. Decisions are based solely on current states and disturbances, prioritizing immediate solar utilization, battery buffering, and grid supplementation when required.

Operational Behavior

As shown in **Figure 6**, RBC operates myopically. Solar energy is allocated instantaneously to the electrolyzer, excess generation charges the battery, and any remaining shortfall is met by grid imports regardless of price. No anticipatory charging or production shifting is observed. The battery acts primarily as a short-term smoothing device, while hydrogen storage is used passively rather than strategically. This behavior reflects purely reactive control without intertemporal coordination.

Energy and Economic Implications

The aggregated performance metrics in **Figure 7** reveal three defining characteristics of RBC operation:

- *High grid dependence*, indicating limited coordination between renewable availability and electrolyzer dispatch.
- *Moderate hydrogen utilization*, suggesting imperfect alignment between production and demand over time.
- *Grid-dominated operating costs*, which substantially constrain net profitability.

Although the controller maintains stable and feasible operation and achieves positive profit, it fails to exploit price variability, perform storage arbitrage, or anticipate future demand shifts. As a result, economic performance remains modest relative to system capacity.

Baseline Significance

RBC establishes a practical lower-bound benchmark for comparison. Its performance illustrates the structural limitations of reactive dispatch under uncertainty and highlights the potential value of predictive optimization and adaptive learning approaches, examined in the following sections.

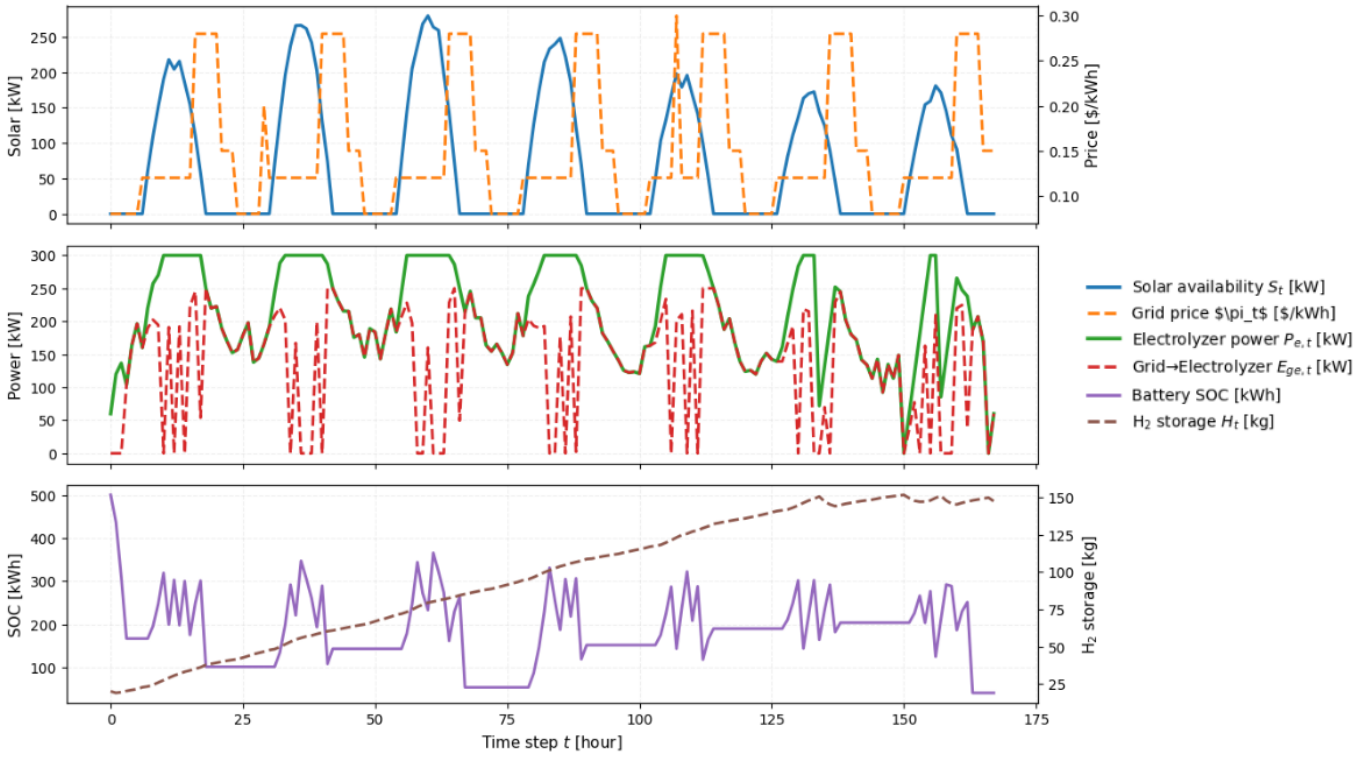


Figure 6. Operational trajectories under rule-based control (RBC), showing solar availability and electricity price (top), electrolyzer and power dispatch (middle), and battery state-of-charge and hydrogen storage levels (bottom).

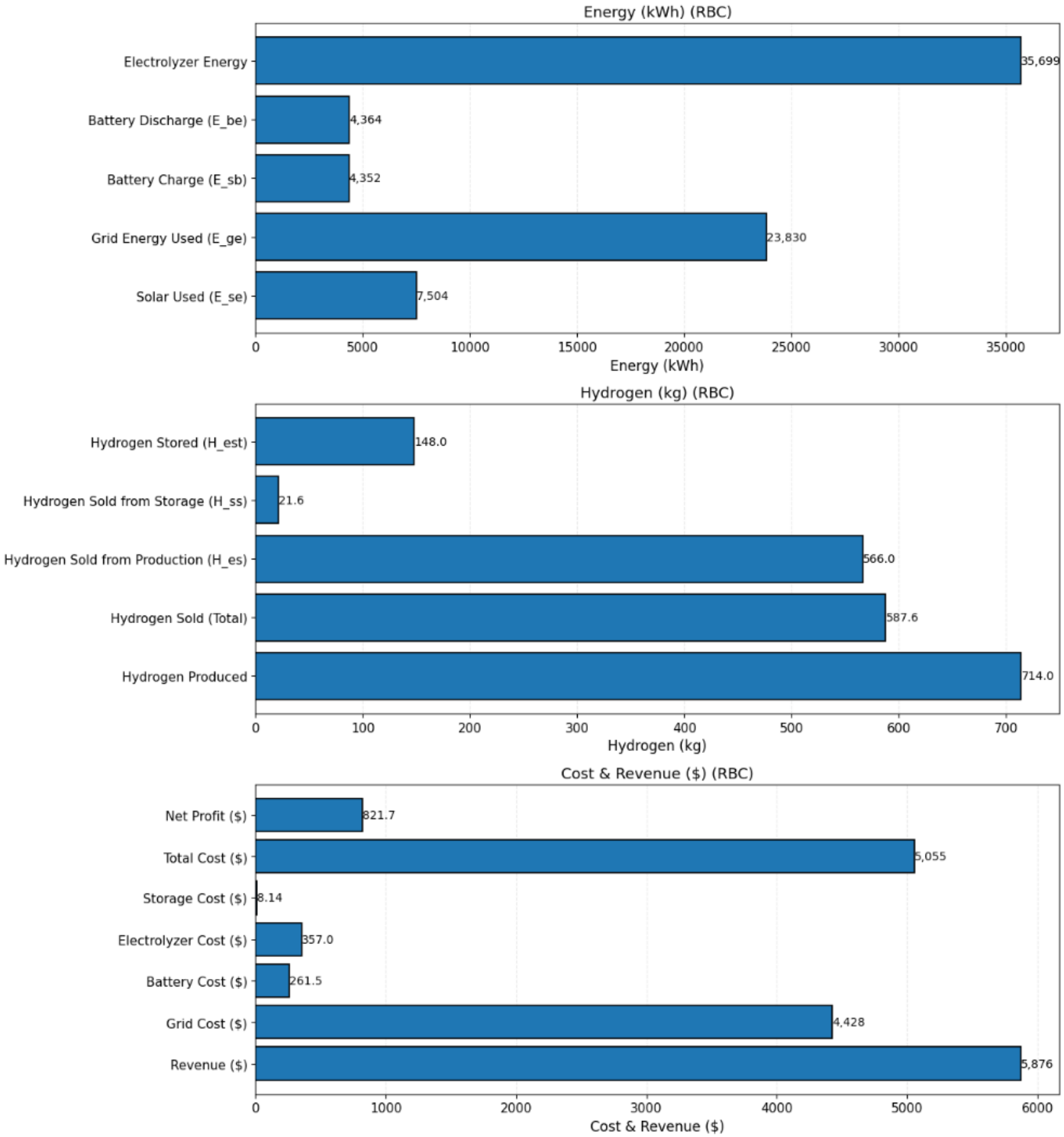


Figure 7. Aggregated performance metrics for rule-based control (RBC), including energy flows, hydrogen production and utilization, and economic cost–revenue breakdown.

Model Predictive Control (MPC)

Operational Behavior

The operational trajectories in **Figure 8** show structured and economically coordinated dispatch behavior. Electrolyzer power remains consistently high, frequently operating near rated capacity. Grid imports are not constant but vary in alignment with price signals, indicating economically rationalized grid usage rather than unconditional supplementation.

Battery operation is deliberate and episodic. Charging occurs during favorable periods, while discharge is deployed strategically to support electrolyzer operation during less favorable renewable conditions. The battery therefore functions as an economic flexibility resource rather than a simple smoothing buffer.

Hydrogen storage exhibits controlled oscillations without persistent accumulation or depletion. Production and withdrawal remain tightly synchronized, reflecting coordinated temporal management rather than reactive balancing. The overall trajectory indicates stable, constraint-respecting operation with clear intertemporal structure.

Energy and Economic Implications

The aggregated metrics in **Figure 9** reveal three defining characteristics of MPC operation:

- Sustained high electrolyzer utilization, indicating effective capacity exploitation.
- Reduced cost intensity relative to throughput, reflecting economically structured grid usage.
- Strong hydrogen production–sales alignment, resulting in near-complete utilization of produced hydrogen.

Revenue remains dominant relative to total operating cost, and storage-related costs are negligible. Grid imports remain present but are deployed in a targeted and economically justified manner. The resulting net profit reflects coordinated dispatch across renewable availability, price variability, and demand realization.

System-Level Significance

MPC demonstrates stable, economically efficient operation under realistic constraints. The structured dispatch behavior observed in Figure 8 directly translates into the improved aggregated performance shown in Figure 9. The controller maintains feasibility while leveraging temporal coordination to enhance overall system profitability.

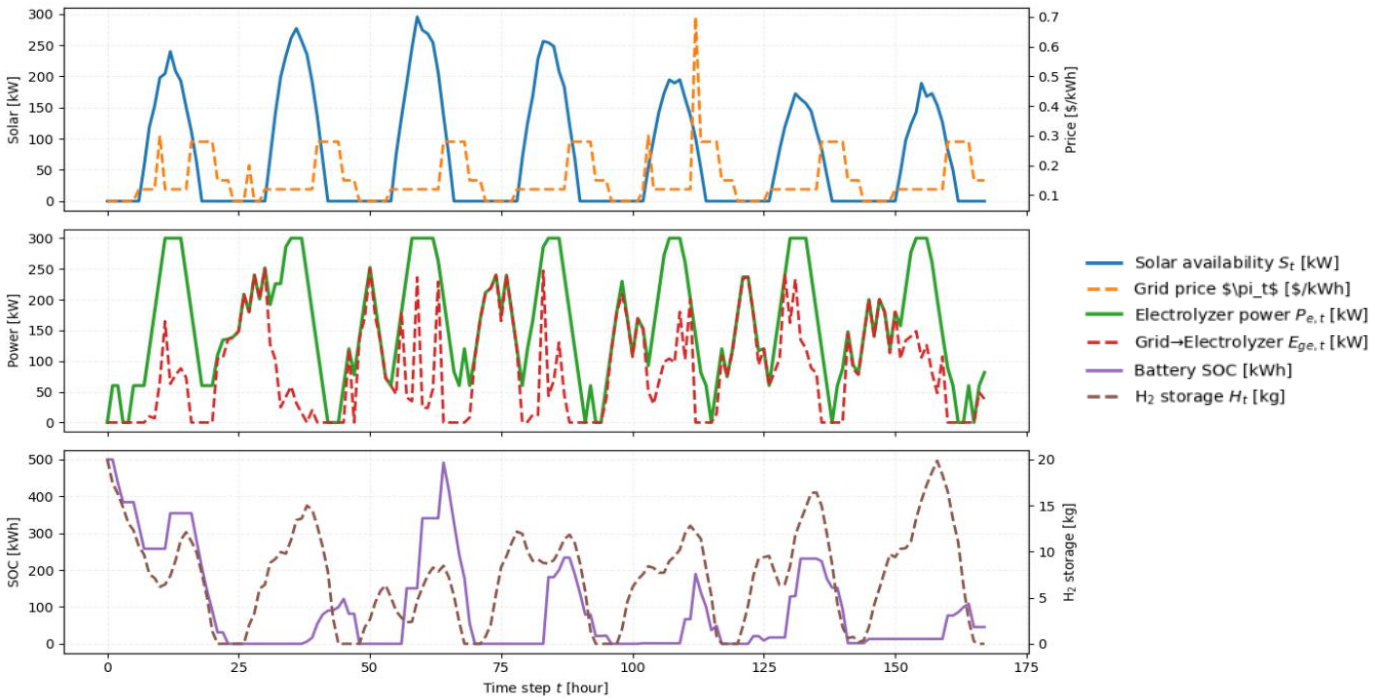


Figure 8. Operational trajectories under model predictive control (MPC), showing solar availability and electricity price (top), electrolyzer and grid dispatch (middle), and battery state-of-charge and hydrogen storage levels (bottom).

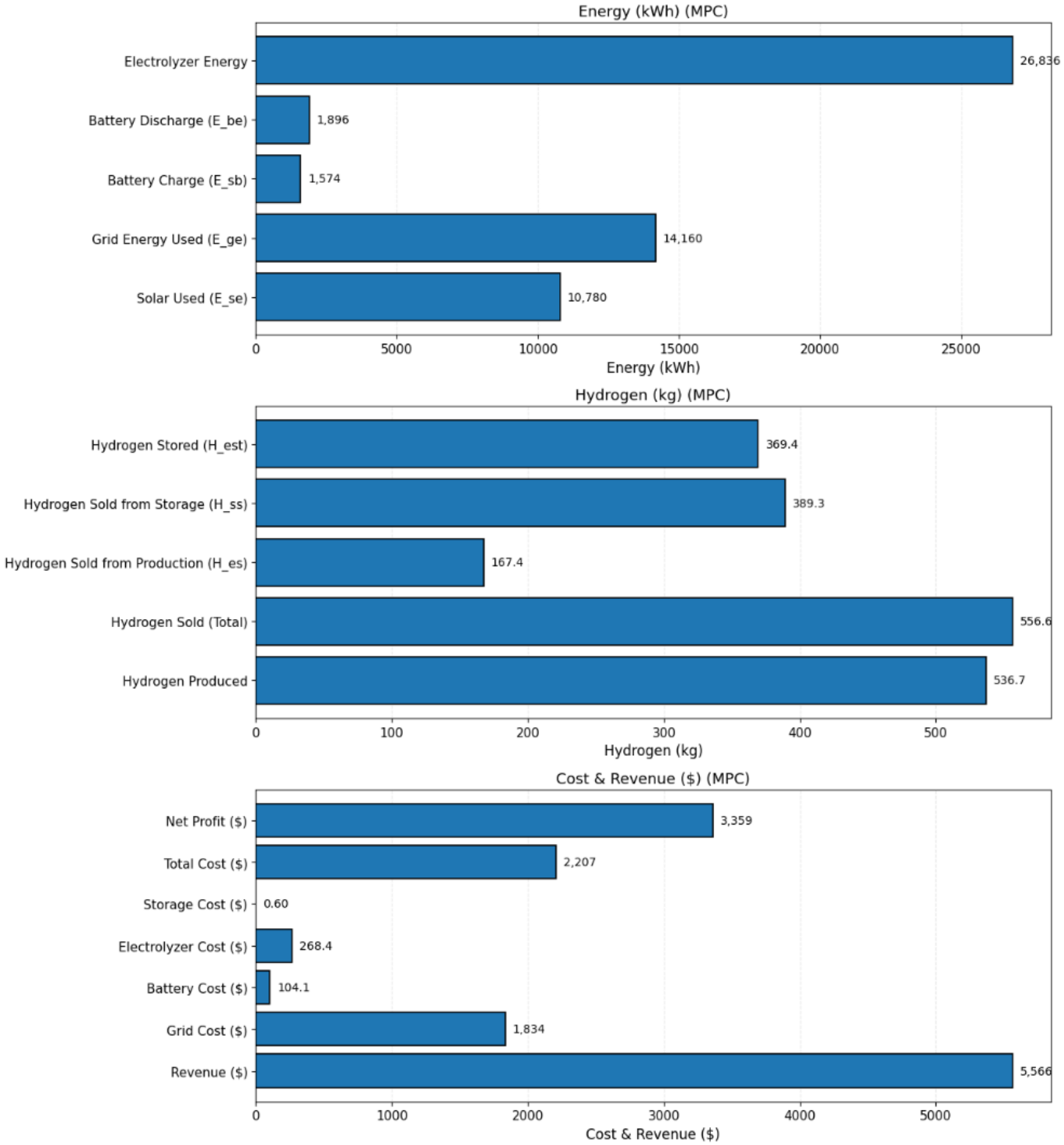


Figure 9.Aggregated performance metrics for model predictive control (MPC), including energy flows, hydrogen production and utilization, and economic cost–revenue breakdown.

Reinforcement Learning without Forecasts (RL-NF)

Operational Behavior

The trajectories in Figure 10 show stable but reactive dispatch under partial observability. Electrolyzer power closely tracks real-time solar availability, increasing during renewable peaks and declining during low-solar periods. Grid imports fluctuate accordingly, supplementing renewable shortfalls without evidence of anticipatory price-based scheduling.

Battery participation is minimal, with negligible charging and limited discharge. As a result, temporal arbitrage is largely absent, and operational flexibility is achieved primarily through direct solar–grid balancing. Hydrogen storage is drawn down early and thereafter remains near low levels, indicating production is aligned closely with immediate demand rather than coordinated intertemporal cycling.

Overall, the controller maintains feasibility and stable operation but lacks the structured temporal coordination observed in predictive control.

Energy and Economic Implications

The aggregated results in **Figure 11** highlight three core characteristics:

• Strong direct solar utilization with moderate grid supplementation.
• Limited battery engagement.
• High production–sales alignment, resulting in near-complete hydrogen utilization.

Revenue clearly exceeds total operating cost, and storage-related costs are negligible. Grid energy represents the dominant expense component. While net profit remains strong, total production scale is lower than under MPC, reflecting the absence of forecast-driven ramping during favorable conditions.

System-Level Significance

RL-NF demonstrates that learning-based control can achieve economically competitive and constraint-respecting operation without explicit forecasts. However, as shown in **Figure 10**, the dispatch pattern remains reactive rather than intertemporally structured. The aggregated performance in **Figure 11** therefore reflects efficient but myopic decision-making, underscoring the value of predictive information for unlocking additional flexibility and profitability.

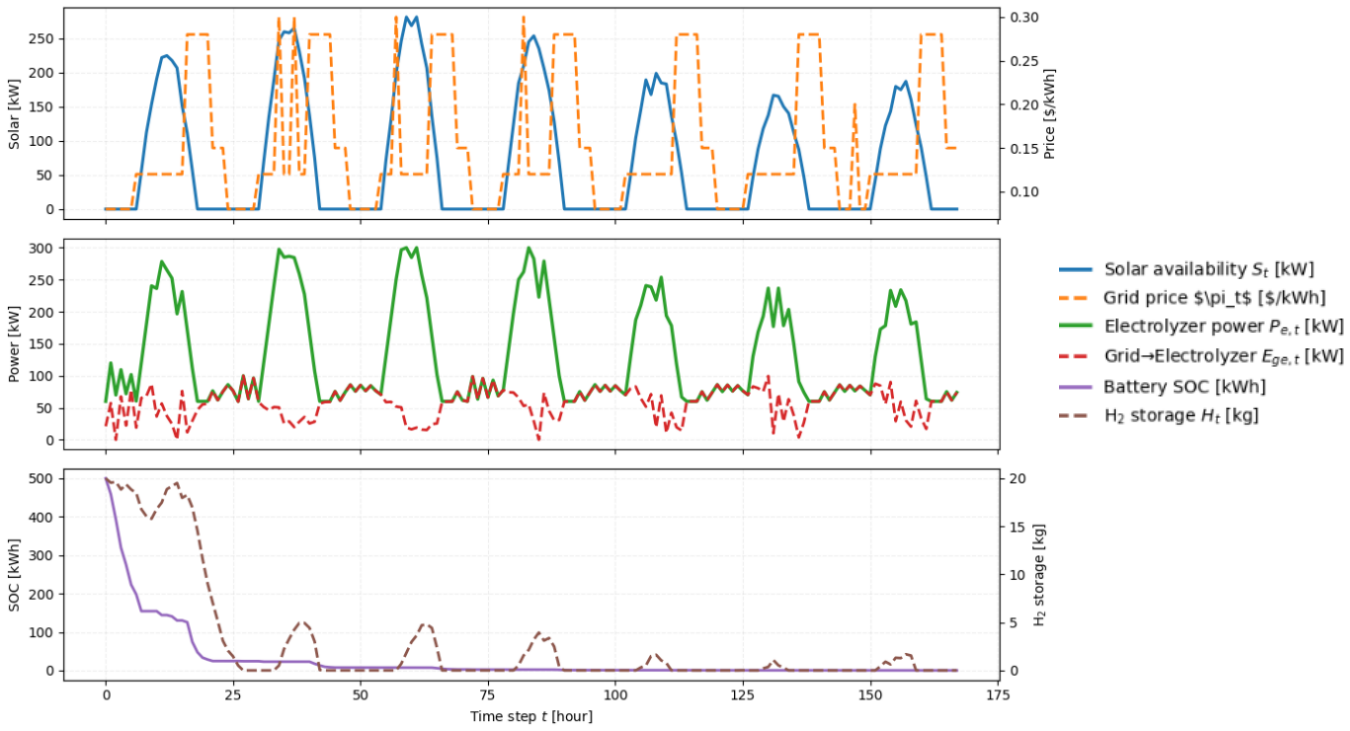


Figure 10. Operational trajectories under reinforcement learning without forecasts (RL-NF), showing solar availability and electricity price (top), electrolyzer and grid dispatch (middle), and battery state-of-charge and hydrogen storage levels (bottom).

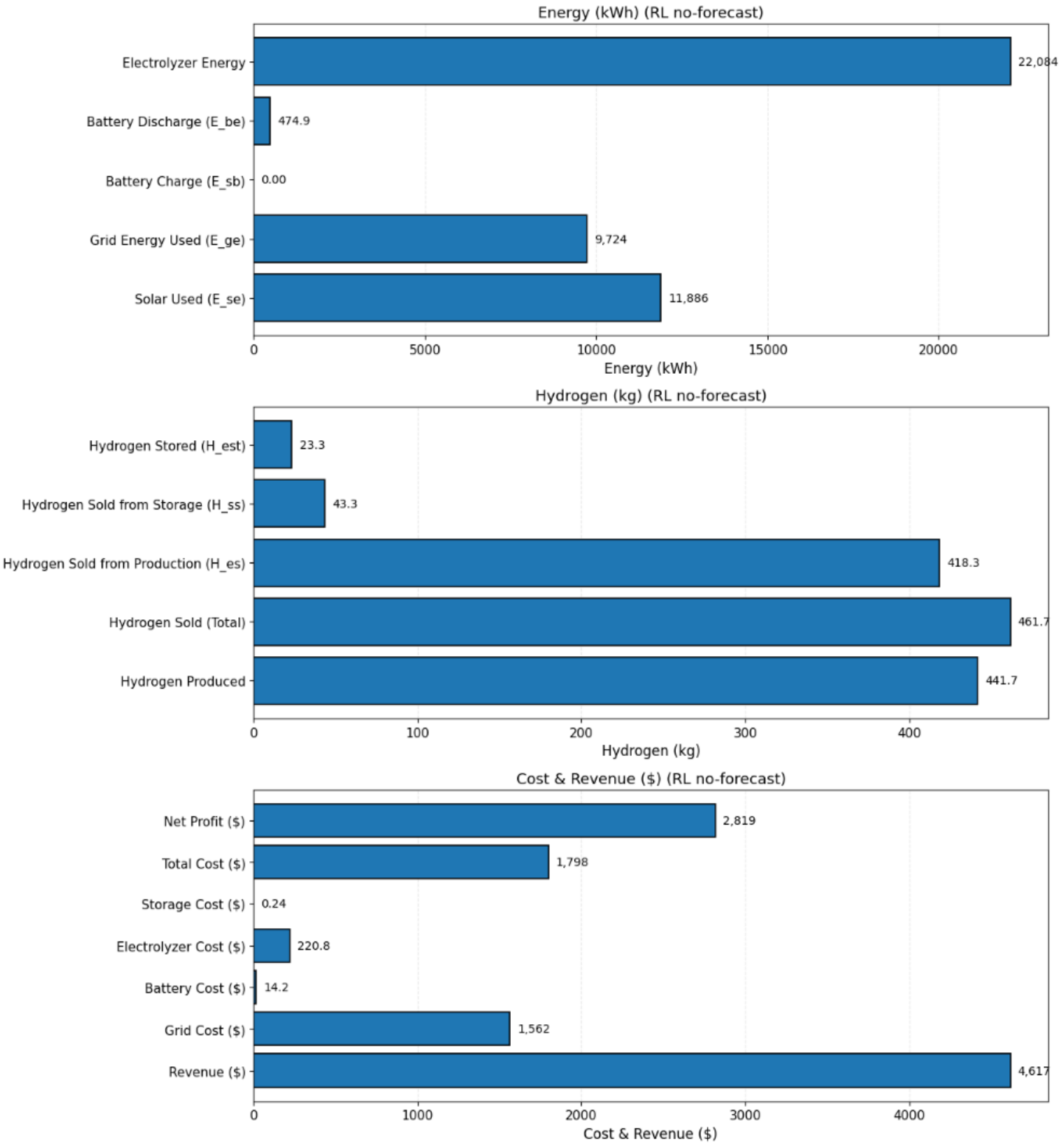


Figure 11. Aggregated performance metrics for reinforcement learning without forecasts (RL-NF), including energy flows, hydrogen production and utilization, and economic cost–revenue breakdown.

Reinforcement Learning with Forecast (RL-F)

Operational Behavior

Figure 12 shows a more structured and forward-looking dispatch compared to the no-forecast case. Electrolyzer power does not simply follow real-time solar availability; instead, operation remains sustained during favorable periods, indicating the use of short-horizon forecasts for anticipatory scheduling.

Grid imports are adjusted in a coordinated manner rather than serving only as immediate compensation. Battery participation remains limited but purposeful. Hydrogen storage is managed more smoothly, avoiding rapid depletion and reflecting improved synchronization between production and expected demand. Overall, the operation remains stable and constraint-respecting, with clearer intertemporal structure.

Energy and Economic Implications

The aggregated results in Figure 13 show:

• Strong renewable utilization with balanced grid support.
• Near-complete hydrogen utilization.
• Negligible storage and battery costs relative to revenue.

Revenue clearly exceeds total cost, and net profit approaches the MPC benchmark. Compared to RL without forecasts, total production is similar; the primary improvement lies in better timing of decisions rather than increased scale.

System-Level Significance

Incorporating forecasts enhances temporal coordination without requiring explicit optimization. The controller shifts from reactive behavior to structured, forward-looking operation, improving economic alignment while maintaining learning-based adaptability.

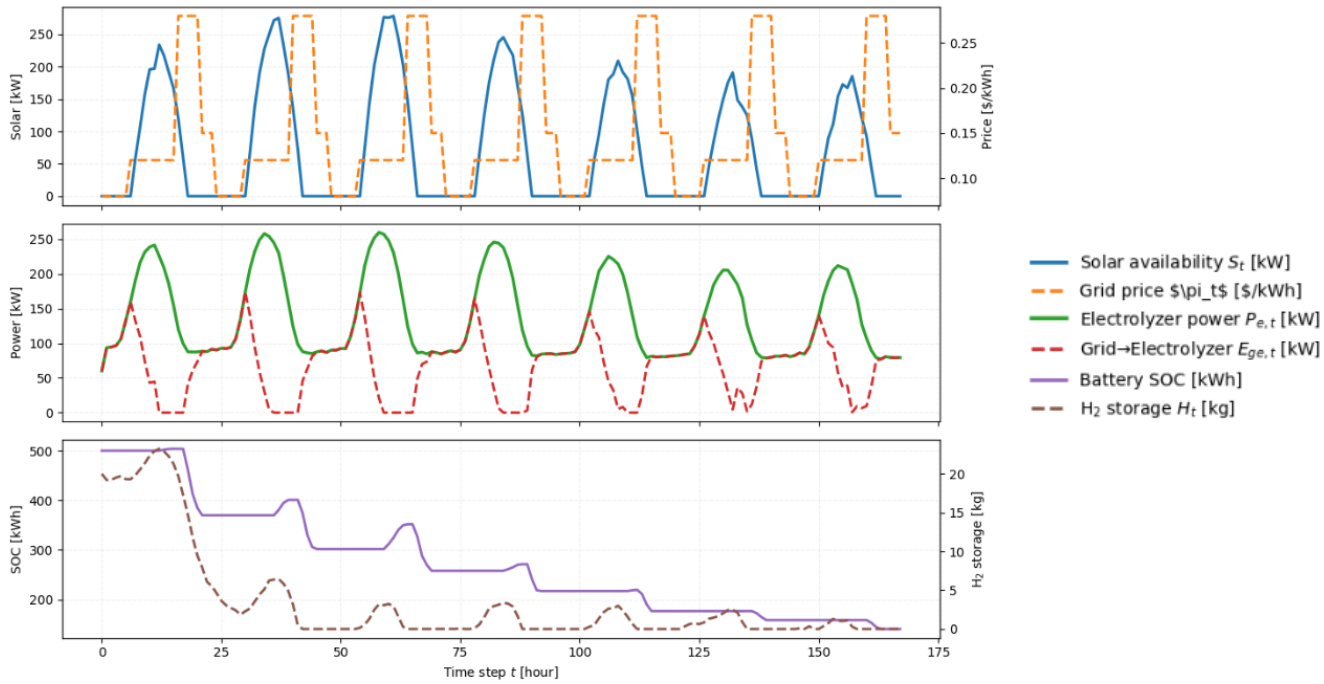


Figure 12. Operational trajectories under reinforcement learning with forecasts (RL-F), showing solar availability and electricity price (top), electrolyzer and grid dispatch (middle), and battery state-of-charge and hydrogen storage levels (bottom).

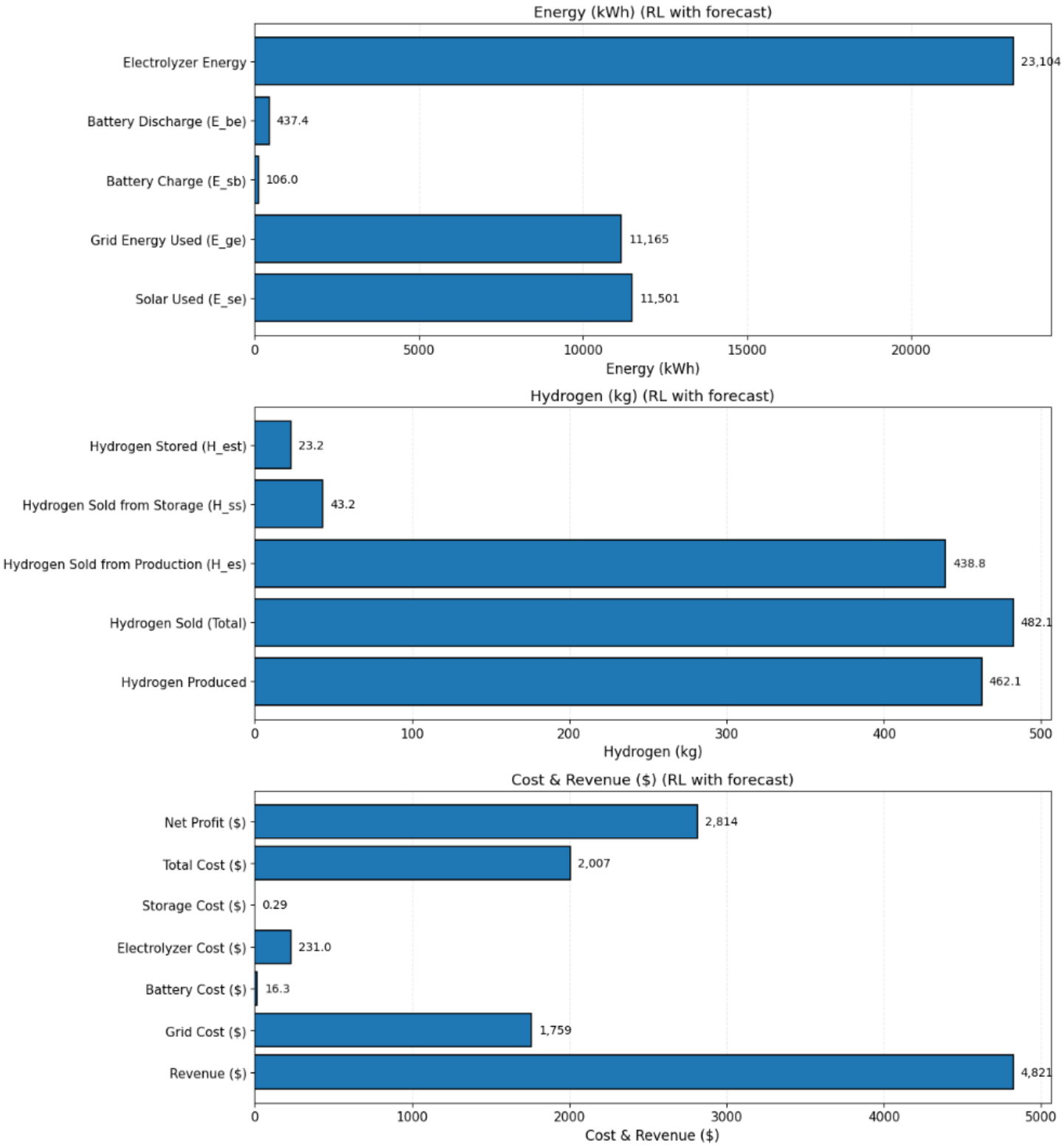


Figure 13. Aggregated performance metrics for reinforcement learning without forecasts (RL-F), including energy flows, hydrogen production and utilization, and economic cost–revenue breakdown.

Conclusion

This study presented a unified and physically consistent comparison of four control strategies for a renewable-powered hydrogen supply chain: rule-based control (RBC), model predictive control (MPC), reinforcement learning without forecasts (RL-NF), and reinforcement learning with forecast-augmented observations (RL-F). All controllers were evaluated under identical disturbances, constraints, and economic assumptions to ensure a transparent and fair comparison.

Three central findings emerge. First, reactive heuristic control (RBC) maintains feasibility and positive profitability but exhibits clear structural inefficiencies, including high grid dependence and reduced hydrogen utilization (~80%). This highlights the economic limitations of myopic dispatch under uncertainty. Second, MPC achieves the strongest economic performance and near-perfect hydrogen utilization by coordinating electrolyzer operation, storage dynamics, and grid interaction through explicit forecast-based optimization. The results demonstrate the value of structured intertemporal planning when predictive information is available. Third, reinforcement learning exhibits competitive and robust behavior, particularly in the no-forecast setting, achieving high utilization and substantial profit improvement over heuristic control. However, incorporating noisy forecasts does not consistently enhance RL performance, indicating that learning-based exploitation of predictive signals remains condition-dependent.

It is important to note that the analysis relies on several simplifying assumptions, including stylized disturbance models, a simplified market structure, and a merchant-operation objective without contractual delivery penalties. Furthermore, RL outcomes are specific to the chosen modeling, training, and forecast configurations. Consequently, while the comparative trends observed here are informative, broader validation under more realistic system representations and multiple uncertainty realizations is necessary to establish fully general conclusions.

Overall, the findings clarify the relative strengths of heuristic, optimization-based, and learning-based control paradigms for hydrogen supply-chain operation under uncertainty, and provide a structured benchmark for future comparative studies.

Future Work

Several extensions can further advance the control of hydrogen supply chains under uncertainty.

Hybrid MPC–RL
Combining MPC's structured, forecast-based optimization with RL's adaptability is a promising direction. Hybrid designs—such as RL-tuned MPC, MPC-guided RL training, or supervisory switching frameworks—could preserve constraint satisfaction and economic efficiency while improving robustness to model mismatch and changing operating conditions.

Robust MPC
Since MPC performance depends on forecast quality, future work should incorporate explicit uncertainty handling. Scenario-based, chance-constrained, or distributionally robust MPC formulations can reduce sensitivity to forecast errors and improve reliability under renewable and demand variability.

Multi-Agent Reinforcement Learning (MARL)
As hydrogen systems scale and decentralize, multi-agent RL can enable coordinated control of distributed assets (electrolyzers, storage units, grid interfaces). This approach supports scalability and modular deployment, though stability, coordination, and safety constraints remain open challenges.

References

Behzadi, A., Alirahmi, S. M., Yu, H., & Sadrizadeh, S. (2023). An efficient renewable hybridization based on hydrogen storage for peak demand reduction: A rule-based energy control and optimization using machine learning techniques. *Journal of Energy Storage, 57*, 106168.

Brka, A., Al-Abdeli, Y. M., & Kothapalli, G. (2016). Predictive power management strategies for stand-alone hydrogen systems: Operational impact. *International Journal of Hydrogen Energy, 41*(16), 6685–6698.

Glavic, M., Fonteneau, R., & Ernst, D. (2017). Reinforcement learning for electric power system decision and control: Past considerations and perspectives. *IFAC-PapersOnLine, 50*(1), 6918–6927.

Huang, C., Zong, Y., You, S., & Træholt, C. (2022). Economic model predictive control for multi-energy system considering hydrogen-thermal-electric dynamics and waste heat recovery of MW-level alkaline electrolyzer. *Energy Conversion and Management, 265*, 115697.

International Energy Agency (IEA). (2019). *The Future of Hydrogen: Seizing Today's Opportunities.* Paris: International Energy Agency.

Liang, T., Chai, L., Cao, X., Tan, J., Jing, Y., & Lv, L. (2024). Real-time optimization of large-scale hydrogen production systems using off-grid renewable energy: Scheduling strategy based on deep reinforcement learning. *Renewable Energy, 224*, 120177.

Oh, E., & Wang, H. (2020). Reinforcement-learning-based energy storage system operation strategies to manage wind power forecast uncertainty. *IEEE Access, 8*, 20965–20976.

Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., & Morari, M. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings, 45*, 15–27.

Parisio, A., Rikos, E., & Glielmo, L. (2014). A model predictive control approach to microgrid operation optimization. *IEEE Transactions on Control Systems Technology, 22*(5), 1813–1827.

Rawlings, J. B., Mayne, D. Q., & Diehl, M. M. (2020). *Model Predictive Control: Theory, Computation, and Design.*

Ruelens, F., Claessens, B. J., Vandael, S., De Schutter, B., Babuška, R., & Belmans, R. (2016). Residential demand response of thermostatically controlled loads using batch reinforcement learning. *IEEE Transactions on Smart Grid, 8*(5), 2149–2159.

Staffell, I., Scamman, D., Abad, A. V., Balcombe, P., Dodds, P. E., Ekins, P., Schmidt, P., Bauen, A., & Ward, K. R. (2019). The role of hydrogen and fuel cells in the global energy system. *Energy & Environmental Science, 12*(2), 463–491.

Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy, 235*, 1072–1089.

Velarde, P., Valverde, L., Maestre, J. M., Ocampo-Martínez, C., & Bordons, C. (2017). On the comparison of stochastic model predictive control strategies applied to a hydrogen-based microgrid. *Journal of Power Sources, 343*, 161–173.

Yang, T., Zhao, L., Li, W., & Zomaya, A. Y. (2021). Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning. *Energy, 235*, 121377.

Ye, J., Wang, X., Hua, Q., & Sun, L. (2024). Deep reinforcement learning based energy management of a hybrid electricity-heat-hydrogen energy system with demand response. *Energy, 305*, 131874.