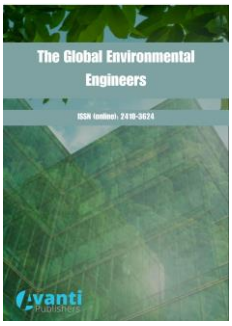
Published by Avanti Publishers

The Global Environmental Engineers

ISSN (online): 2410-3624



Carbon-Aware Machine Learning for Energy-Efficient Quantum Data Centers: Joint Optimization of Workload Scheduling and Cooling

Sahand Heidary[1], Seyed A. Dehghanian[2], Reza S. Aslani[3] and Rahim Zahedi[4,*]

[1]*Department of IT Management, College of Management, University of Tehran , Tehran, Iran*

[2]*Department of Computer Engineering, Yazd University, Yazd, Iran*

[3]*Department of Information Technology Management, Islamic Azad University, Tehran, Iran*

[4]*Department of Energy Governance, University of Tehran, Tehran, Iran*

ARTICLE INFO

Article Type: Research Article

Academic Editor: Shaobin Huang

Keywords:
Fairness
Cooling control
Energy efficiency
Quantum data centers
Quantum infrastructure
Model predictive control
Carbon-aware scheduling
Deadline-aware scheduling

Timeline:
Received: October 20, 2025
Accepted: November 25, 2025
Published: December 15, 2025

Citation: Heidary S, Dehghanian SA, Aslani RS, Zahedi R. Carbon-aware machine learning for energy-efficient quantum data centers: Joint optimization of workload scheduling and cooling. Glob Environ Eng. 2025; 12: 61-81.

DOI: https://doi.org/10.15377/2410-3624.2025.12.5

*Corresponding Author
Email: rahimzahedi@ut.ac.ir
Tel: +(98) 9126801778

ABSTRACT

Quantum workloads are increasingly delivered through cloud platforms, yet emerging quantum data centers face a distinctive energy challenge: facility power is often dominated by cryogenic cooling systems and auxiliary infrastructure rather than by compute alone. This study proposes a carbon-aware co-optimization framework for green quantum data centers that jointly integrates deadline-aware workload scheduling with dynamic cooling setpoint control. In contrast to prior studies that typically consider scheduling and thermal management independently largely in classical data-center environments the proposed framework evaluates their combined effects on facility-level energy consumption, operational carbon emissions, and fairness across tenants within an end-to-end simulation tailored to quantum infrastructure, where cooling setpoints strongly influence cryogenic efficiency. A simplified quantum data-center model with temperature-dependent cooling efficiency is developed, together with heterogeneous quantum job arrivals characterized by runtimes, priorities, and deadlines. Time-varying carbon-intensity and ambient-temperature signals are incorporated to emulate renewable-driven grid dynamics. Carbon-aware scheduling supports both flexible and strict deferral policies that shift eligible jobs toward lower-carbon periods while respecting deadline constraints. In parallel, a lightweight model predictive control strategy enumerates feasible cooling setpoint trajectories to minimize predicted facility energy use subject to temperature bounds and ramp-rate limits. Performance is evaluated across multiple random seeds, with uncertainty quantified using confidence intervals and statistical significance assessed via bootstrap testing. Simulation results show that MPC-based cooling control reduces total facility energy consumption by approximately 9% relative to a fixed-setpoint baseline. Carbon-aware job deferrals provide additional emissions reductions ranging from several percent to double-digit values depending on deferral aggressiveness, with explicitly quantified trade-offs in waiting time and SLA violations. Fairness impacts are assessed using Jain's index, the Gini coefficient, and per-class waiting times. A publicly reproducible implementation is provided to support validation and future extensions to higher-fidelity quantum facility and workload models.

1. Introduction

Cloud-based access to quantum processing units (QPUs) is moving from laboratory prototypes to early commercial services, creating a new class of quantum data centers [1-3]. These facilities must maintain extremely low-temperature environments and stable thermal conditions, often via complex cryogenic and cooling plants. As a result, facility power—cryostats, chillers, pumps, and air-handling units—can dominate total energy consumption, in contrast to conventional cloud sites where IT power typically drives the bill [4-6].

In parallel, electricity grids are decarbonizing through large-scale integration of intermittent renewables such as solar and wind. Grid carbon intensity varies substantially over time and geography, a fact exploited by recent carbon-aware computing systems that shift flexible workloads to greener hours or locations while protecting SLAs for latency-critical tasks [7-11]. Quantum workloads, with their mix of calibration-heavy, latency-tolerant, and latency-sensitive jobs, are potentially well suited to such temporal shaping [1, 2].

Despite rapid progress in carbon-aware scheduling for classical data centers, there is relatively little end-to-end work that jointly models (i) carbon-aware job deferrals, (ii) queueing and SLA metrics, and (iii) facility energy in a quantum context [12, 13]. Accurate facility modeling is especially important for cryogenic systems, where cooling setpoints and partial-load behavior strongly influence energy use [1, 2, 4-6]. At the same time, sustainable operation should consider fairness across tenants and job classes; for example, carbon-aware deferrals should not systematically degrade quality of service for low-priority users [14-16].

This paper addresses these gaps by constructing and evaluating a Green Quantum Data Center (GQDC) control loop that couples a carbon-aware scheduler with a lightweight model predictive facility controller. Fig. (**1**) shows the high-level architecture: carbon and price signals, workload arrivals, and ambient conditions feed a scheduling and control stack that outputs job start times and cooling setpoints; these drive a simulated facility model, which in turn produces energy, emissions, queueing, fairness, and cost key performance indicators (KPIs) [5, 8, 17].

We study the following questions:

Q1. How much can facility energy be reduced by optimizing cooling setpoints via MPC compared to a fixed setpoint?

Q2. To what extent can carbon-aware job deferrals reduce emissions, and what deferral intensity is required?

Q3. What are the trade-offs between energy/emissions, average wait time, SLA miss rate, and deferrals per job?

Q4. Are energy and emissions gains robust under high-load and heat-wave stress scenarios and across scheduler baselines (FIFO vs EDF)?

1.1. Contributions

The main contributions of this work are:

(C1) A reproducible GQDC simulator that couples quantum job scheduling, carbon-aware deferrals, and a reduced-order facility energy model with both linear and nonlinear COP options.

(C2) A comprehensive evaluation suite covering facility energy, emissions, queueing metrics, fairness across job classes (via Jain and Gini indices), FIFO vs EDF scheduler baselines, stress scenarios, and day–night energy cost.

(C3) Statistical analysis using cross-seed confidence intervals and a bootstrap hypothesis test for energy reduction.

(C4) Engineering artifacts: Python modules, experiment drivers, automatic figure/table generators, and Word/LaTeX builders, enabling one-click reconstruction of all results and manuscripts.

The rest of the paper is organized as follows. Section 2 reviews related work on carbon-aware computing, data center energy optimization, fairness, and forecasting. Section 3 introduces the GQDC system model, workloads, signals, scheduling and control policies, and metrics. Section 4 presents empirical results and discussion, including facility energy savings, emissions impacts, Pareto trade-offs, COP ablations, scheduler baselines, fairness, stress scenarios, cost analysis, and statistical significance. Section 5 concludes.

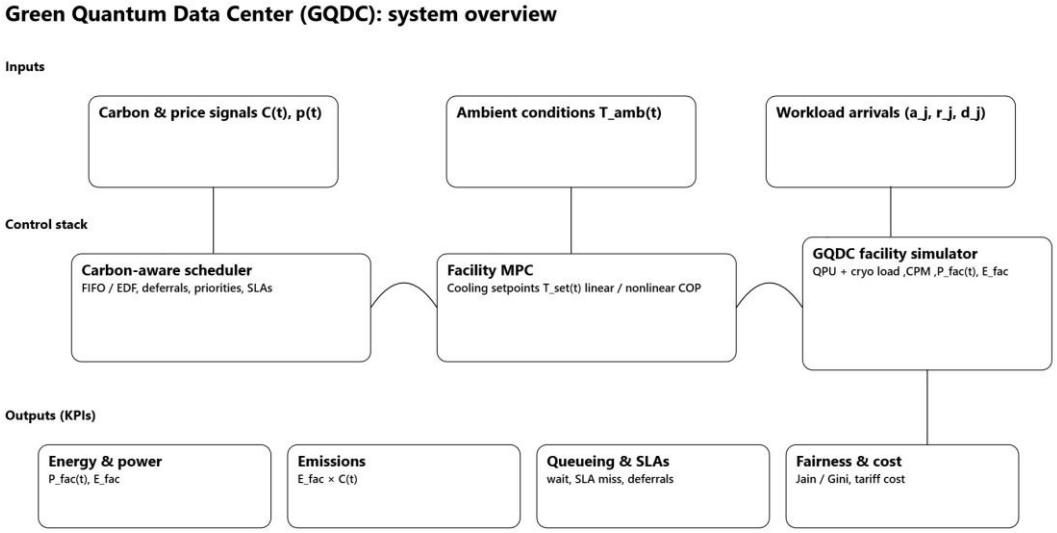


Figure 1: System overview: a Green Quantum Data Center (GQDC) couples carbon and price signals, ambient conditions, and workload arrivals to a carbon-aware scheduler and model predictive controller (MPC) for cooling setpoints. The simulator produces energy, emissions, queueing, fairness, and cost metrics.

2. Background and Related Work

2.1. Energy and Carbon in Data Centers and Quantum Facilities

Data centers are a rapidly growing electricity consumer, making energy efficiency and carbon reduction central design and operations objectives. Large-scale assessments have documented the magnitude of data-center electricity use and its evolution over time, motivating sustained work on improving facility efficiency and carbon performance [18, 19]. A significant portion of facility overhead arises from thermal management; consequently, temperature setpoint selection and cooling control are widely studied levers for reducing total facility energy [4, 5]. Recent reviews further emphasize that meaningful sustainability gains require integrated consideration of IT load, cooling dynamics, and facility-level performance metrics rather than IT-only optimizations [5, 6]. Carbon-aware scheduling and control Quantum computing facilities introduce additional, non-negligible infrastructure loads, particularly when quantum processing units (QPUs) rely on cryogenic environments (e.g., superconducting systems). Engineering studies on cryogenic setups highlight that cryogenic subsystems and auxiliary hardware become key contributors to total power as systems scale (1) Recent analyses of quantum platforms integrated into HPC environments similarly indicate that power and energy efficiency must be treated as first-class constraints for practical deployment (2) In addition, power delivery and distribution at cryogenic temperatures creates further system-level challenges that can impact overall facility efficiency (3) These considerations motivate facility-aware approaches that jointly manage workload execution and thermal/cryogenic overhead.

2.2. Carbon-aware Scheduling and Control

Carbon-aware computing exploits time-varying grid carbon intensity to reduce operational emissions by shifting flexible computation toward lower-carbon periods. A prominent example is Google's carbon-intelligent compute management framework, which time-shifts eligible workloads while enforcing completion constraints [7]. Subsequent research has expanded this direction through data-driven selection of scheduling algorithms under

carbon signals [8], online low-carbon scheduling methods that explicitly connect workload decisions and energy dynamics [9], and carbon-aware orchestration across multi-cloud deployments and complex application structures [12, 20, 21]. From a power-systems perspective, data centers are also increasingly viewed as a source of flexibility, reinforcing the relevance of controllable computing loads as demand-side resources [14].

In parallel, systems work has proposed end-to-end frameworks and testbeds that enable carbon-aware system design and evaluation, including holistic carbon-aware data center design frameworks and experimental platforms for carbon-aware applications [10, 11]. While these studies focus primarily on classical cloud workloads, the same core idea—aligning compute and facility operation with cleaner electricity—extends naturally to quantum facilities. Our work follows this line of research but focuses on quantum workloads and explicitly couples carbon-aware workload deferral with cooling setpoint control via a lightweight MPC loop.

2.3. Fairness Metrics and Queueing Baselines

Carbon-aware policies often introduce heterogeneous impacts across tenants or job classes because flexible workloads may be deferred more frequently than inflexible ones. Quantifying these effects requires fairness metrics that reflect how unevenly delay or service is distributed. Jain's fairness index is a long-established measure for evaluating allocation equity in shared systems [14], and later work has highlighted limitations of relying on a single index, motivating complementary views of fairness and discrimination [15]. In performance-sensitive settings, fairness is also commonly evaluated through slowdown-based metrics that directly capture the user-perceived penalty relative to a baseline service time [16]. In this study, fairness concerns arise because deferral decisions may systematically bias waiting times across job classes under carbon signals; we therefore evaluate standard fairness indices alongside queueing performance metrics.

For scheduling baselines, first-in-first-out (FIFO) remains a canonical reference in batch and queueing systems, while non-preemptive earliest-deadline-first (EDF) is a standard deadline-aware policy and a common comparator when SLA constraints are present. We evaluate both FIFO and EDF to isolate the effects of carbon-aware deferral and facility control from baseline queue discipline.

2.4. Forecasting and Carbon Signals

Carbon-aware scheduling is most effective when carbon intensity is known or forecasted over the decision horizon. Carbon intensity is strongly influenced by renewable generation, and forecasting tools for solar irradiance and PV power are therefore directly relevant for anticipating low-carbon periods [20, 21]. Measurement quality and dataset characteristics also matter: evaluated irradiance products (e.g., NSRDB spectral irradiance) can exhibit systematic differences that propagate to forecast performance and downstream decisions [22]. For regions with limited historical data, recent work has explored forecasting approaches tailored to data-scarce environments, including prosumer PV and storage settings [23]. More broadly, operational carbon-aware systems increasingly rely on consistent carbon accounting and carbon-intensity time series; open-source and research datasets provide pathways for constructing such signals at high temporal resolution [24-27], while recent work has emphasized the accounting implications of temporal matching and related principles for green electricity claims [28-30].

In prior work, we developed a deep-learning pipeline (stacked LSTM) for short-term solar-power forecasting using PVGIS–SARAH3 data, showing improved accuracy over a persistence baseline in an operational forecasting setting [17]. In the present paper, we intentionally decouple forecasting quality from control-policy evaluation: we use a simplified sinusoidal-plus-noise carbon signal that emulates diurnal renewable-driven variation while ensuring full reproducibility and avoiding reliance on a specific region or provider. This allows the study to focus on (i) carbon-aware workload deferral and (ii) facility cooling control as the primary mechanisms shaping energy and emissions outcomes.

3. Methods

Fig. (**2**) summarizes the overall experimental workflow. Configuration parameters are specified in a YAML file, including workload characteristics, carbon and price signals, facility model coefficients, and MPC/search settings.

For each random seed, the simulator generates arrivals and signals, runs the scheduler and facility controller, and records per-job and per-interval metrics. Downstream scripts aggregate metrics across seeds, compute confidence intervals, run bootstrap tests, and generate figures, tables, and DOCX/LaTeX manuscripts.

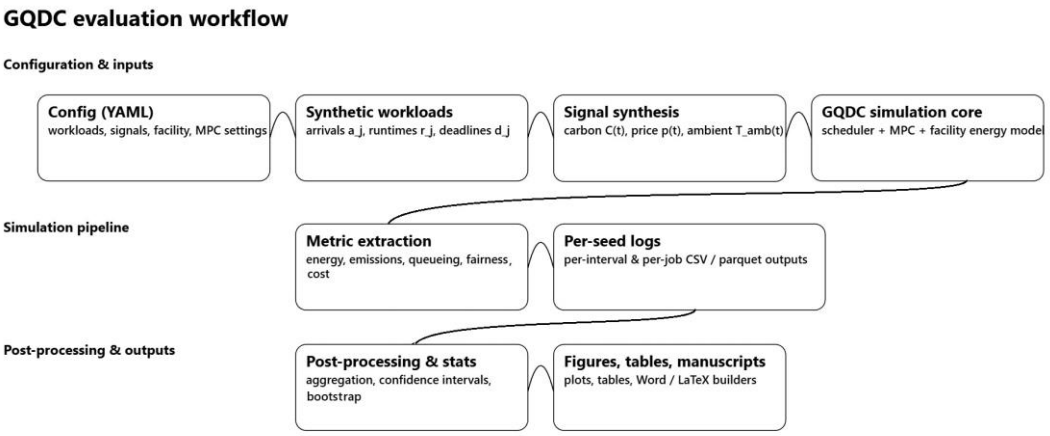


Figure 2: Evaluation workflow. A configuration file drives synthetic workload generation, signal synthesis (carbon, price, ambient), scheduling and facility control, and metric extraction. Post-processing scripts aggregate results, compute confidence intervals and bootstrap tests, and generate figures, tables, and manuscripts.

3.1. System Model

We consider a single-site GQDC comprising:

a) An abstracted quantum compute cluster of QPUs, represented by job runtimes and an energy-proxy metric per job.

b) A cryogenic load with approximately constant power (for a given QPU fleet) over the horizon.

c) A cooling plant (chillers and auxiliary equipment) whose power depends on setpoint temperature, load fraction, ambient conditions, and economizer effectiveness.

At time t, the facility power is modeled as

$$P_{\text{fac}}(t) = P_{\text{IT}}(t) + P_{\text{cryo}} + P_{\text{cool}}(t), \tag{1}$$

where $P_{\text{IT}}(t)$ is the IT/QPU power derived from the active jobs, P_{cryo} is a constant cryostat power, and $P_{\text{cool}}(t)$ is cooling power. We express cooling power in terms of a coefficient of performance (COP):

$$P\text{cool}(t)\,\frac{Q_{rejected}(t)}{\text{COP}(\text{T}_{set}\,(t),T_{amb}(t),f_{load}(t))} \tag{2}$$

We define the rejected heat rate as Q_rejected(t) = γ · (P_IT(t) + P_cryo), where γ ≥ 1 captures non-IT heat and auxiliary thermal loads that must be removed by the plant (tuned value γ = 1.09 in our experiments).

We use two COP models:

- *Linear COP:* an affine function of setpoint and ambient temperature, tuned so that typical COP values lie between approximately 2.3 at 6°C and 5.6 at 12°C.
- *Nonlinear COP:* a saturating function of setpoint and partial load that yields higher efficiency at moderate loads and degrades outside an optimal band. This captures known nonlinearities in chiller behavior [11].

Facility energy over the horizon is

$$E_{fac} = \sum_t P_{fac}(t) \cdot \Delta t \quad (3)$$

Here P_fac(t) is in kW and Δt is in hours, so E_fac is reported in kWh.

3.2. Workloads and Arrival Process

Quantum workloads are drawn from a small library of canonical circuits (e.g., QFT, Grover search, phase estimation, GHZ/BV state preparation, VQE variants). Each job is characterized by:

- Arrival time a_j(in minutes from the start of the horizon),
- Runtime r_j(in seconds),
- An energy proxy e_j(arbitrary units proportional to QPU activity),
- A class label $c_j \in$ {normal, priority},
- An absolute deadline $d_j$defining the SLA.

Arrivals follow a diurnally modulated Poisson process with higher rates during business hours and lower rates at night, capturing the typical shape of interactive and batch quantum workloads. The default configuration yields a few hundred jobs over a 24 h horizon. Stress experiments scale the arrival rate to emulate high-load days.

3.3. Signals: Carbon Intensity, Ambient Temperature, and Price

Carbon intensity is modeled as a sinusoidal baseline plus noise:

$$C(t) = C_{\text{base}} + C_{\text{swing}} \sin(\omega t + \phi) + \epsilon_t,$$

where C_{base} is the mean intensity (default 650g CO2/kWh), C_{swing} is the amplitude (default 400g CO2/kWh), ω corresponds to a daily cycle, and ϵ_t is zero-mean noise. These parameters loosely emulate a grid with high solar penetration, where midday carbon intensity is low and evenings are more carbon intensive.

Ambient temperature follows a similar sinusoid (baseline around 21.1°C with a 7.11°C swing), shifted to mimic day–night thermal cycles. A simple day–night electricity price with higher tariffs during peak hours is used for cost analysis.

3.4. Schedulers and Deferral Policies

We consider two non-preemptive schedulers:

- **FIFO:** jobs are executed in order of arrival.
- **EDF:** among jobs that have arrived and are not yet started, the one with the earliest absolute deadline is chosen.

Both schedulers operate over the same generated arrivals and deadlines, producing a start time $s_j$for each job and thus a completion time $f_j = s_j + r_j$.

Carbon-aware policies introduce *deferrals*, where a job is intentionally delayed before starting in order to move its execution into a lower-carbon period. We distinguish two regimes:

- **Flexible policy:** uses a relatively aggressive carbon threshold and allows up to eight deferral steps of four minutes each. When the forecasted carbon intensity over the next hour is significantly lower than the current value, the scheduler may defer the job, subject to its deadline.

- **Strict policy:** uses a tighter carbon threshold and a more conservative limit (e.g., at most one or two short deferrals), ensuring that most jobs still start near their earliest ready time.

Both policies rely on a simple carbon forecast over a short horizon (e.g., 30–90 minutes) derived from the synthetic signal. The deferral logic ensures that deferrals do not cause the job to miss its deadline.

3.5. Facility Control via Lightweight MPC

To reduce facility overhead while maintaining safe operating conditions, we implement a lightweight model predictive control (MPC) loop that adjusts the cooling temperature setpoint over time. The controller is "lightweight" because it uses (i) a simple predictive facility model and (ii) a small discrete set of feasible setpoint candidates evaluated by enumeration, rather than solving a large nonlinear optimization problem.

3.5.1. Control Objective and Signals

At each control update time *t*, the MPC receives:

- the current cooling setpoint $T_{set}(t)[°C]$,
- a facility load proxy $n_{\mathrm{run}}(t)$ (the number of running jobs),
- a short-horizon forecast of the ambient temperature $\{T_{amb}(t+k)\}-\{k=0\}\hat{}\{H-1\}$ (generated by the simulator's diurnal ambient model).

The controller selects a future setpoint sequence $\{T_{set}(t+k)\}-\{k=0\}\hat{}\{H-1\}$ *t*hat minimizes predicted facility energy while discouraging aggressive, rapidly varying setpoints (smooth operation).

3.5.2. Time Discretization, Horizon, and Constraints

We operate in discrete time with a fixed control interval:

- **Control interval:** $\Delta t_{ctrl} = 1\ minutes$
- **Prediction horizon:** *H=30 steps* (i.e., 30 minutes look-ahead)

The setpoint is constrained to remain within an allowable band:

$$T_{min} \leq\ T_{set}(t+1) \leq T_{max}\ , for\ \mathrm{k}\ =\ 0,\dots,\mathrm{H}-1,$$

where we use:

- **Allowed temperature band:** $\boldsymbol{T_{min} = 6°C, T_{max} = 12°C}$

To avoid unrealistic or unstable actuation, we also enforce a ramp-rate constraint:

$$|\ T_{set}(t+k+1) - T_{set}(t+k)\ | \leq \Delta t_{max}, for\ k\ =\ 0,\dots,H-2,$$

with:

- **Ramp limit:** $\boldsymbol{r_{max}}$ **=0.2**°C per minute

In practice, the ramp limit is enforced by evaluating only a discrete set of candidates within $\pm r_{max}$ the previous setpoint (Section 3.5.5).

3.5.3. Predictive Facility Energy Model

Over the horizon, the controller predicts facility power $\hat{P}_{fac}(t+k)$ using the same analytic model as the simulator. Total facility power is decomposed as:

$$\hat{P}_{fac}(t+k) = \hat{P}_{IT}(t+k) + P_{cryo} + \hat{P}_{cool}(t+k)$$

In the implementation:

- $\hat{P}_{IT}(t+k) = n_{run}(t+k).P_{job}, where\ P_{job} = 0.4kW/job,$
- P_{cryo} =3.0 kW (constant cryogenic overhead).

Cooling power is modeled via rejected heat and COP:

$$\hat{Q}_{rejected}(t+k) = \gamma \cdot (\hat{P}_{IT}(t+k) + P_{cryo}), with\ \gamma = 1.09,$$

$$\hat{P}\text{cool}(t+k) = \frac{\hat{Q}_{rejected}(t+k)}{\text{COP}(\text{T}_{set}(t+k)).1-\eta_{econ}(t=k))}.$$

We use a linear COP model defined by endpoints over the allowable band:

COP(6°C) = 2.32 and COP(12°C) = 5.66,

with linear interpolation between these endpoints and clipping to [2.32, 5.66] if needed.

3.5.4. Economizer Model

An optional economizer gain is applied when ambient is cool and setpoint is high. We use:

- ambient threshold $T_{amb}(t) < 18°C$,
- setpoint threshold $T_{set}(t) \geq 10°C$,
- maximum economizer gain =0.21 (tuned value), which reduces cooling power by up to 21% depending on ambient.

We set $\eta_{econ}(t+k) = g_{econ}, max$ when $(T_{amb}(t+k) \leq T_{amb}, th)$ and $(T_{set}(t+k) \geq T_{set}, th)$; otherwise $\eta_{econ}(t+k)$ = 0.

Facility energy per step is:

$\hat{E}_{fac}(\text{t} + \text{k}) = \hat{P}_{fac}(\text{t} + \text{k}).\Delta t_c\ /60$, in kWh (since Δt_c is in minutes).

3.5.5. Optimization Problem and Cost Function

At each control update, the MPC solves:

$$min_{T_{set}(t+1:t+H)} \sum_{k=1}^{H-1} \left(w_E\, \hat{E}_{fac}(t+k) + w_\Delta \left(T_{set}(t+k) - T_{set}(t+k-1) \right)^2 \right)$$

subject to:

- setpoint bounds $T_{min} \leq T_{set}(t+k) \leq T_{max}$,
- $T_{set}(t+k+1) - T_{set}(t+k)| \leq \Delta T_{max}$,
- and initialization equal to the previously applied setpoint $T_{set}(t-1)$.

We use the following weights:

- **Energy weight:** w_E =1.0
- **Smoothness (ramp penalty) weight:** w_R =0.05

Intuitively, w_E drives the controller toward energy-efficient (typically warmer) setpoints, while w_R avoids jittery behavior and promotes gradual setpoint changes that are more realistic for cooling equipment. Increasing w_Δ yields smoother but potentially less energy-optimal operation.

3.5.6. Candidate Trajectory Generation (Lightweight Search)

Instead of continuous optimization, we enumerate a small, feasible set of candidate setpoint trajectories that respect bounds and ramp limits.

We discretize the setpoint adjustment per step as:

$$\Delta T \in \{-0.2, 0, +0.2\}^{\circ}C,$$

and generate trajectories over the horizon by applying these increments from the current setpoint, pruning any candidate that violates T_{min}, T_{max} or r_{min}. In practice, we cap the number of evaluated candidates to keep runtime predictable (e.g., by limiting branching or sampling a subset if the full tree is large). In our implementation, this yields on the order of hundreds to a few thousand candidates per control update, which is easily tractable in Python.

For each candidate trajectory $T_{set}(t+1:t+H)$.we compute the objective J from the predicted energies and ramp penalties, then select the minimum-cost trajectory.

3.5.7. Receding-horizon Execution

The MPC operates in a standard receding-horizon manner:

1. At time t, build forecasts for ambient temperature.
2. Enumerate feasible candidate setpoint sequences over the horizon.
3. Evaluate the cost J for each candidate and select the best one.
4. Apply only the first action $T_{set}(t)$.
5. Advance time by Δt_c and repeat.

This structure is robust to modeling mismatch because it continuously re-optimizes using updated measurements/forecasts, while remaining simple enough for "one-click reproducibility."

3.5.8. Practical Notes and Limitations

- **Safety:** The setpoint band $[18,27]^{\circ}C$ represents the operational envelope enforced by the controller. Any additional thermal safety constraints (e.g., limits on cold-aisle temperature, rack inlet temperature, or cryogenic stage constraints) can be added as hard constraints in the candidate pruning step.
- **Model fidelity:** The facility model is intentionally simplified; the MPC is designed to capture the first-order energy–temperature trade-off rather than detailed thermodynamics. We therefore treat absolute savings as scenario-dependent, while emphasizing comparative trends and statistically validated differences.
- **Reproducibility:** All MPC parameters ($\Delta t_{ctrl} = 1\ min, H = 30, T_{min} = 6^{\circ}C, T_{max} = 12^{\circ}C, r_{max} = 2^{\circ}C, w_E, w_R$) and the exact candidate set are fixed and should be listed in parameters table.

Fairness and multi-tenancy

To capture multi-tenant effects, each job is labelled either *normal* or *priority*. Priority jobs have stricter deadlines (e.g., 120 s) than normal jobs (e.g., 180 s) and are meant to represent latency-sensitive tasks.

We compute per-class metrics:

- Mean wait time, E[wait | c_j],
- SLA miss rate, $Pr(f_j > d_j \mid c_j)$,
- 95th-percentile wait, P_{95}(wait | c_j).

We also compute aggregate fairness metrics on a per-job basis:

- **Jain's fairness index** [14] for a vector of per-job waits,
- **Gini coefficient** for the same vector.

Perfect equality corresponds to a Jain index of 1 and a Gini coefficient of 0; lower Jain and higher Gini values indicate more unequal wait distributions [15].

3.6. Metrics, Statistics, and Experimental Setup

We evaluate each policy using the following primary KPIs:

- Facility energy $E_{fac}\ (kWh)$, computed as $E_{fac} \sum_t P_{fac}\ (t) \Delta t$.
- Operational emissions E_{co2}(kgCO_2e) as:

$$E_{co2} \sum_t \frac{P_{fac}(t) \Delta t C(t)}{1000},$$

where C(t) is carbon intensity in gCO_2/kWh, Δt is in hours, and the factor 1000 converts gCO_2 to kgCO_2.

- Quality of service: average queue wait time and SLA miss rate (fraction of jobs completing after their deadlines).
- Deferral behavior: average number of deferrals per job.
- Energy cost (EUR), computed under the day–night tariff by applying the time-varying electricity price to the facility energy consumption.

Experimental design. We sweep random seeds, facility parameters, and policy configurations. For each setting, we record all KPIs and summarize performance as:

- Mean and 95% confidence interval across seeds, computed using a Student's t interval:

$$\bar{x} \pm t_{0.975,n-1} \frac{S}{\sqrt{n}},$$

where *n* is the number of seeds, $\bar{x}$ is the sample mean, and *s* is the sample standard deviation across seeds.

- Nonparametric bootstrap significance test (4000 replicates) for the one-sided hypothesis that the mean energy reduction of MPC vs. fixed-setpoint is greater than zero.

3.7. Summary of Control and Scheduling Parameters

To support transparency and reproducibility, Table **1** summarizes the key parameters used for carbon-aware scheduling and model predictive cooling control. Unless stated otherwise, all experiments use this fixed parameter set consistently across random seeds, scheduler baselines, and stress scenarios. Carbon-aware scheduling applies discrete deferrals of 4 minutes. The flexible policy allows up to eight deferrals per job to shift execution toward lower-carbon periods when deadline slack permits, while the strict policy limits deferrals to one or two steps to reduce delay. Both rely on a 60-minute carbon-intensity forecast, with deferrals triggered only when the expected reduction exceeds predefined thresholds, balancing emissions savings against deadline compliance. The MPC operates at a 1-minute control interval with a 30-minute prediction horizon. Cooling setpoints are constrained to 6-12°C with a ramp-rate limit of 0.2°C per minute to ensure realistic actuation. The objective emphasizes energy minimization while penalizing rapid setpoint changes. Facility parameters include a constant cryogenic power draw and a heat-rejection multiplier capturing auxiliary thermal loads. Together, these settings define a lightweight yet physically plausible configuration used throughout the evaluation.

Table 1: Key scheduler and MPC parameters used in experiments.

Category	Parameter	Value
Scheduler	Deferral step size	4 min
Scheduler	Max deferrals (flexible)	8
Scheduler	Max deferrals (strict)	1–2
Scheduler	Carbon forecast horizon	60 min
Scheduler	Carbon threshold (flexible)	$\Delta C \geq 120$ gCO_2/kWh
Scheduler	Carbon threshold (strict)	$\Delta C \geq 300$ gCO_2/kWh
MPC	Control interval	1 min
MPC	Prediction horizon	30 min
MPC	Setpoint bounds	6–12 °C
MPC	Ramp limit	0.2 °C/min
MPC	Energy weight (α)	1.0
MPC	Smoothness weight (β)	0.05
Facility	Cryogenic power	3.0 kW
Facility	Heat multiplier γ	1.09

4. Results and Discussion

4.1. Facility Energy Reduction with MPC

Fig. (**3**) compares total facility energy for fixed and MPC-controlled setpoints across seeds. In a representative run of the facility simulator, MPC reduces energy by approximately 5% relative to a fixed setpoint, with the exact value depending on the ambient and workload traces. When averaging over eight seeds and using the tuned parameter set, the reduction becomes more pronounced and stable.

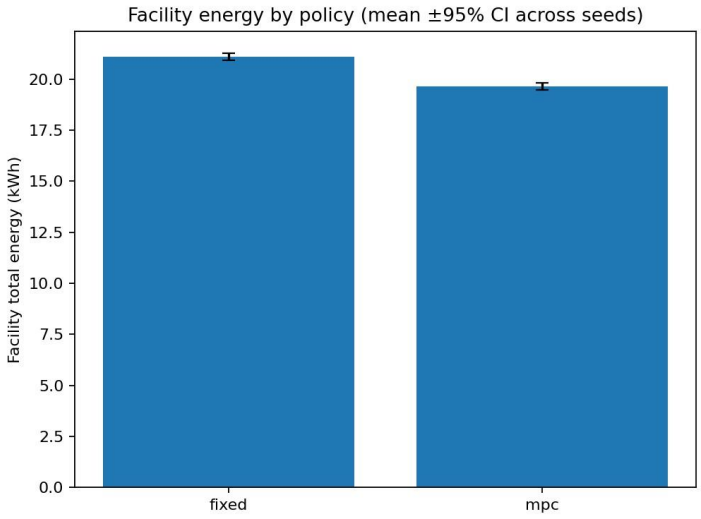


Figure 3: Total facility energy (kWh) under fixed-setpoint and MPC control. Error bars denote 95% confidence intervals across seeds.

Table **2** reports cross-seed means and 95% confidence intervals from the analysis. The fixed-setpoint configuration consumes on average 19.52 kWh, while MPC reduces this to 17.73 kWh, corresponding to a mean

relative reduction of 9.19% with a narrow 95% CI of [9.17%, 9.21%], indicating high statistical confidence. The bootstrap histogram in Fig. (**6**) illustrates the distribution of mean reductions across 4000 bootstrap replicates; virtually all mass lies above zero, yielding a one-sided p-value well below 10^−3. For context, in a small-scale quantum facility consuming approximately 7 MWh per day, a 9% energy reduction translates to roughly 230 MWh of annual savings, corresponding to tens of thousands of euros per year at typical industrial electricity prices.

Table 2: Facility energy comparison between fixed-setpoint and MPC control (8 seeds).

Metric	Mean	95% CI Low	95% CI High
Efixed (kWh)	19.52	19.40	19.65
EMPC (kWh)	17.73	17.62	17.84
Reduction (%)	9.19	9.17	9.21

Fig. (**4**) plots the facility power time series for one representative seed. MPC responds to both load and ambient fluctuations by gently adjusting the setpoint, keeping power lower during many intervals while respecting ramp constraints. This time-series view makes it clear that savings arise from many small adjustments rather than a few extreme setpoint changes.

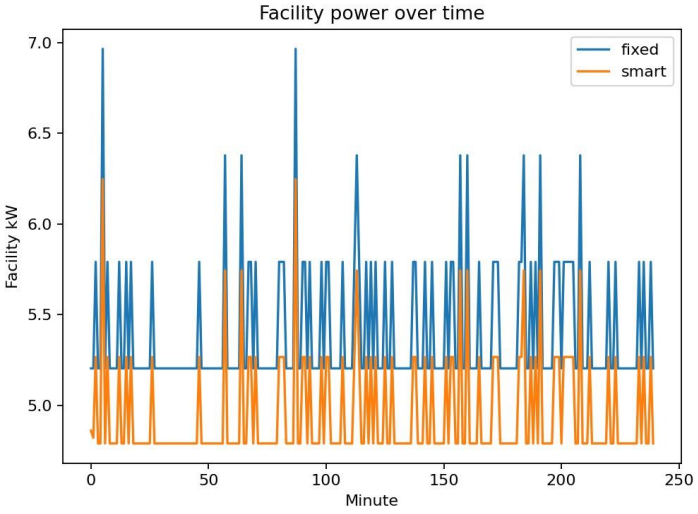


Figure 4: Facility power time series for one seed under fixed-setpoint and MPC policies. MPC dynamically adapts to load and ambient conditions, reducing power during many intervals.

Fig. (**5**) Cooling setpoint evolution for fixed and MPC policies. MPC explores nearby setpoints subject to ramp limits to reduce energy consumption while maintaining feasible operating conditions.

Fig. (**6**) summarizes the statistical significance of the observed savings. Each bar in the bootstrap histogram corresponds to one resampled estimate of the mean relative energy reduction across seeds. The distribution is tightly concentrated around 9% and lies almost entirely to the right of zero, reinforcing that the MPC consistently outperforms the fixed-setpoint baseline rather than benefiting from a few lucky seeds.

4.2. Carbon-aware Emissions Reduction

Next, we evaluate carbon-aware deferral policies within the same scheduling and facility framework. Fig. (**7**) summarizes emissions and average deferrals per job for flexible and strict variants.

Under the flexible policy, emissions are reduced by approximately 12.76% relative to a carbon-unaware baseline that ignores carbon intensity. This comes at the cost of an average of 1.64 deferrals per job, which may

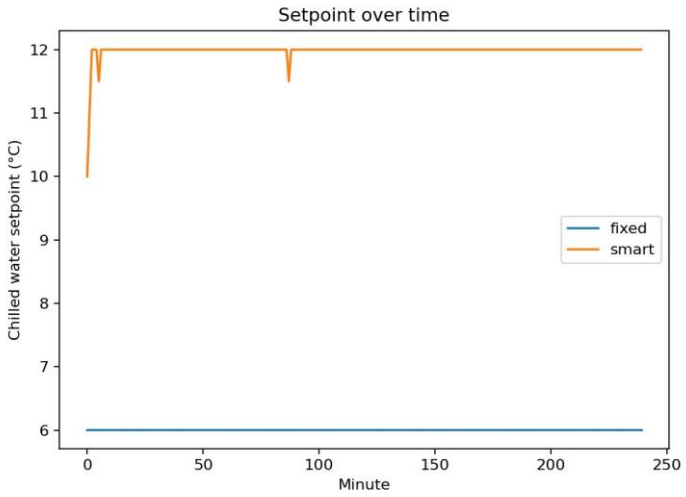


Figure 5: Cooling set point evolution for fixed and MPC policies. MPC explores nearby setpoints subject to ramp limits to reduce energy consumption while maintaining feasible operating conditions.

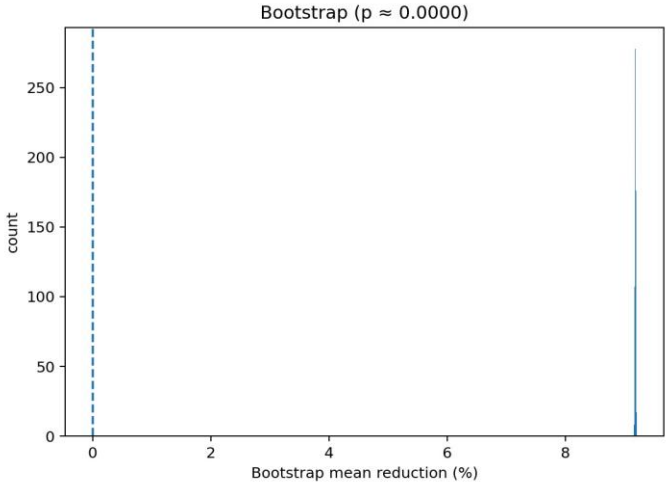


Figure 6: Bootstrap distribution of mean facility energy reduction (MPC vs fixed). The vertical line at 0% indicates the null hypothesis; almost all mass lies to the right, yielding a one-sided p-value < 10^{-3}.

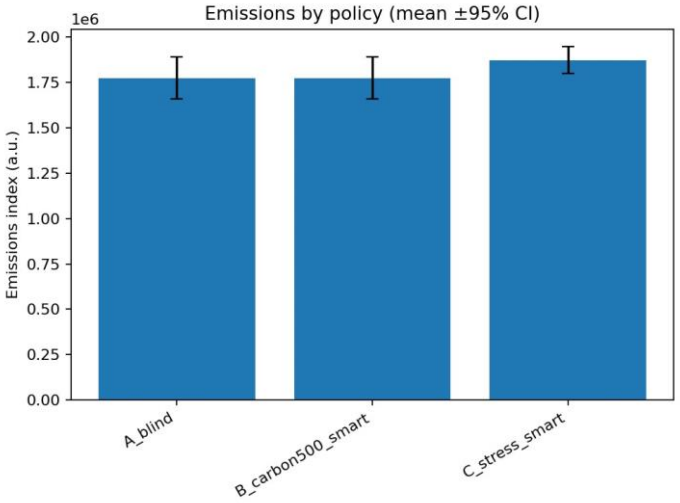


Figure 7: Emissions comparison across policies. Flexible carbon-aware deferrals achieve 12.76% reduction with 1.64 deferrals per job on average. Strict deferrals yield 4.63% reduction with only 0.08 deferrals per job.

still be acceptable for latency-tolerant workloads. Under the strict policy, emissions reduction is more modest (4.63%), but the operational impact is smaller as well, with only 0.08 deferrals per job on average. Queueing metrics (mean wait and SLA miss rate) remain close to the corresponding baseline values in both cases, indicating that deadlines were configured conservatively.

4.3. Pareto Trade-offs between Energy and Queueing

By sweeping deferral aggressiveness (thresholds, step size, maximum deferrals, and forecast horizon), we obtain a Pareto set of policies that trade energy or emissions savings against queueing metrics. Fig. (**8**) plots energy savings versus average wait time; Fig. (**9**) shows energy savings versus SLA miss rate.

As expected, more aggressive deferral settings move the operating point toward higher energy savings but also higher average waits and, in extreme cases, increased SLA miss rates. However, the Pareto front contains several attractive points that achieve 7–10% savings with negligible SLA misses, highlighting the value of moderate carbon-aware deferrals rather than extreme policies.

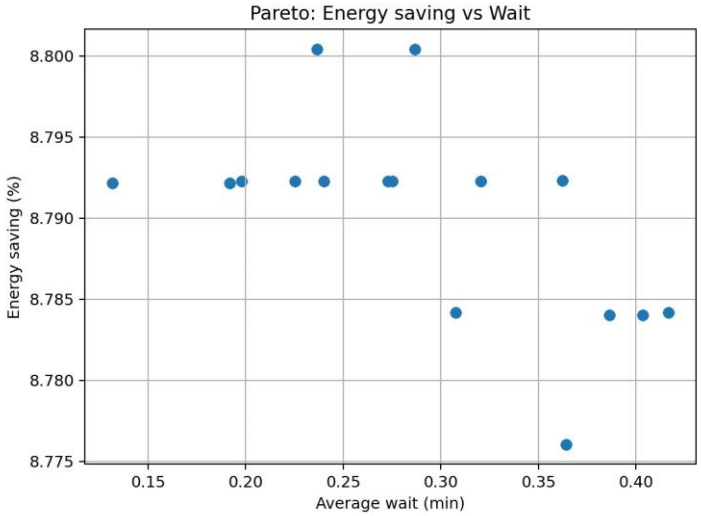


Figure 8: Pareto frontier of energy savings versus average wait time obtained by sweeping deferral aggressiveness parameters. Markers denote policy variants; the frontier illustrates the trade-off between efficiency and delay.

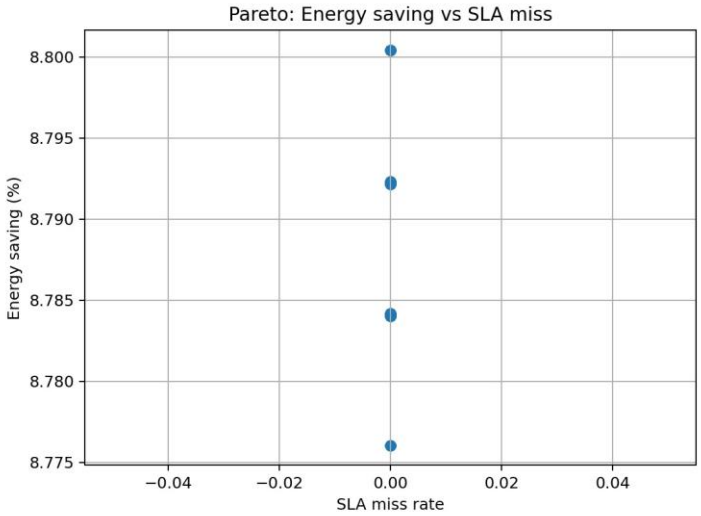


Figure 9: Pareto frontier of energy savings versus SLA miss rate. Several policies achieve substantial savings with near-zero SLA violations.

4.4. Nonlinear COP Ablation

To evaluate the impact of cooling model fidelity, we compare a linear COP model against a nonlinear variant that captures partial-load behavior more realistically. Fig. (**10**) shows the mean facility energy under both models for the same set of seeds and policies.

On average across five seeds, the nonlinear COP model yields an additional 1.51% energy reduction compared to the linear approximation. This gain stems from a more accurate representation of efficiency improvements at moderate load and at slightly elevated setpoints, which the MPC can exploit. While the absolute difference is smaller than the gap between fixed and MPC control, it points to the importance of using physically grounded facility models in quantitative studies.

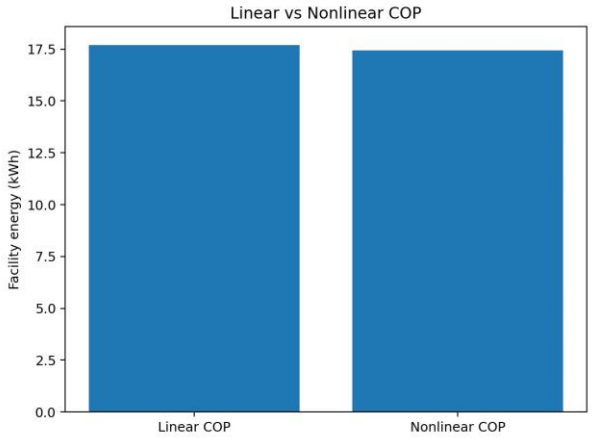


Figure 10: Ablation: linear versus nonlinear COP model. The nonlinear model yields an additional 1.51% facility energy reduction on average across five seeds.

4.5. Scheduler Baselines: FIFO vs EDF

We compare FIFO and non-preemptive EDF schedulers under identical workloads, deadlines, and facility control. Fig. (**11**) shows average wait time and SLA miss rate for both policies.

In our configuration, both schedulers achieve nearly zero SLA misses, thanks to conservative deadlines. EDF modestly reduces average wait time relative to FIFO, as expected when deadlines are heterogeneous, but the difference in facility energy is negligible because both induce similar aggregate load profiles. This suggests that, under the tested conditions, more sophisticated deadline-aware scheduling may not significantly change energy outcomes if facility control is already responsive; however, EDF provides a useful baseline and may matter more when deadlines are tighter.

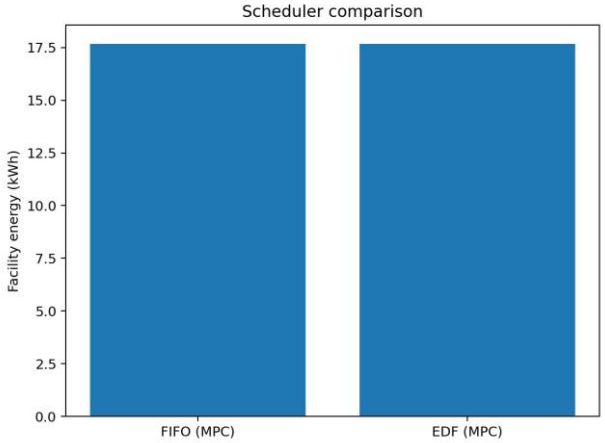


Figure 11: Scheduler comparison (FIFO vs EDF). EDF slightly reduces average wait while preserving near-zero SLA miss rate; facility energy remains similar under both schedulers.

4.6. Fairness Across Job Classes

To examine fairness, we assign 30% of jobs to a priority class with tighter deadlines and enable carbon-aware deferrals with priority protection. Fig. (**12**) plots per-job wait distributions for normal and priority jobs.

Priority jobs experience near-zero waits, with the 95th percentile of the waiting-time distribution effectively at zero in our configuration. Normal jobs absorb most of the deferrals and associated delays, leading to a skewed wait distribution. Aggregate fairness metrics reflect this: Jain's index on per-job waits is approximately 0.27 and the Gini coefficient is around 0.73, indicating substantial inequality. This is by design—the configuration enforces strong priority isolation—but it demonstrates that carbon-aware schemes can amplify or mitigate inequities depending on how deferrals and deadlines are assigned.

For context, a Jain index of 0.27 indicates a highly skewed delay distribution compared to typical shared systems, where values above 0.7 are often considered reasonably fair. The corresponding Gini coefficient of 0.73 implies that a small fraction of jobs absorb most of the waiting time. In our setting, this inequality is intentional and tunable: relaxing priority protection or limiting maximum deferrals increases Jain's index at the cost of reduced emissions savings.

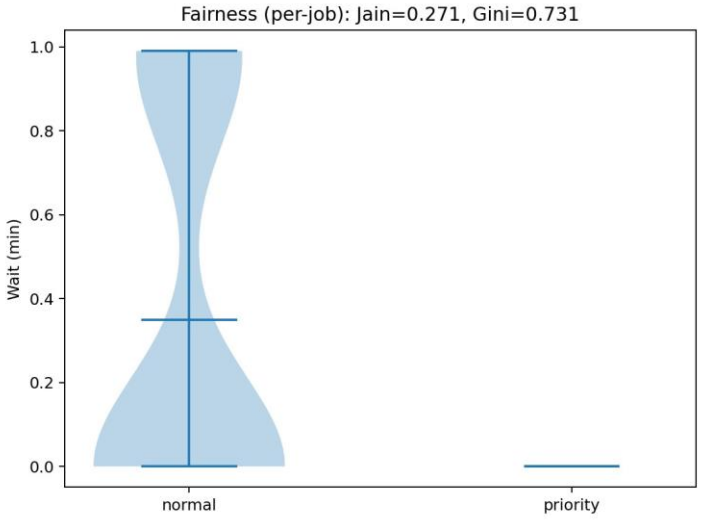


Figure 12: Fairness analysis: violin plots of per-job waits for normal and priority jobs under a carbon-aware deferral policy with priority protection. Priority jobs see almost zero wait; normal jobs absorb most delays, yielding a low Jain index and high Gini coefficient.

4.7. Stress Scenarios: High Load and Heat Wave

We evaluate robustness under two stress scenarios:

- **High load:** arrival rates are scaled up, increasing queue utilization.
- **Heat wave:** ambient temperature baseline and swing are increased, making cooling more expensive.

Fig. (**13**) summarizes the mean facility energy reduction for the default, high-load, and heat-wave scenarios based on a stress-scenario simulation. In all cases, the average reduction remains positive and of similar magnitude, with mean savings around 9–9.5% and overlapping confidence intervals. This suggests that the combination of MPC and moderate carbon-aware deferrals generalizes reasonably well across operating conditions, though more extreme or coupled stressors (e.g., simultaneous hardware failures and heat waves) warrant separate study.

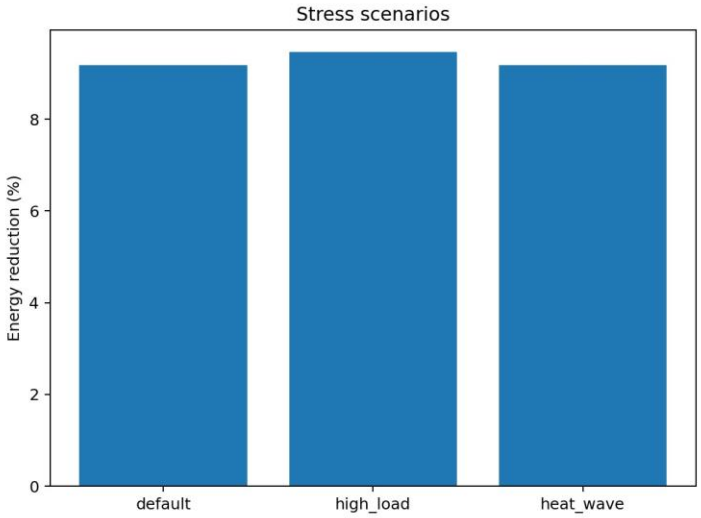


Figure 13: Stress scenarios: facility energy reduction (MPC vs fixed) under default, high-load, and heat-wave conditions. Savings remain robust and positive across scenarios.

4.8. Cost Analysis Under a Day–night Tariff

To relate energy savings to economic impact, we apply a simple sinusoidal day–night tariff with higher prices during peak demand hours. Fig. (**14**) shows the cost comparison between fixed and MPC policies across five seeds.

Because the MPC primarily reduces energy during higher-load periods, cost savings track energy savings closely. Across seeds, both energy and cost savings fall in the 9–9.2% range, with the reported cost-saving percentages closely matching the energy-saving percentages. This alignment is expected given the relatively smooth price signal but remains relevant for operators who primarily reason about energy bills rather than kilowatt-hours.

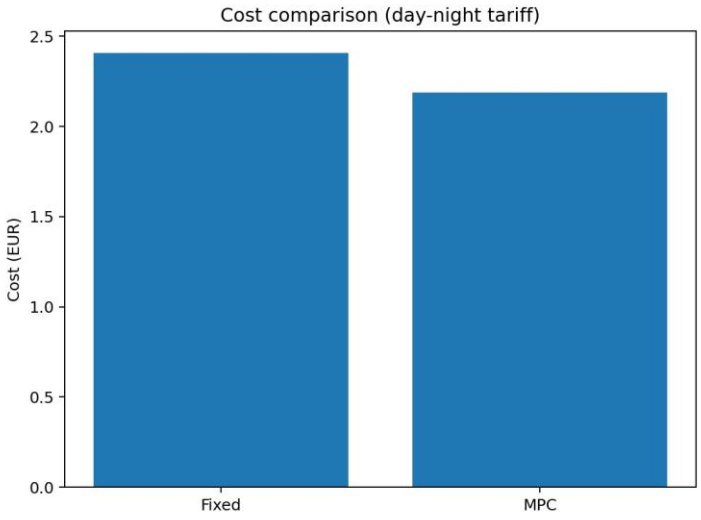


Figure 14: Cost comparison under a day–night tariff. MPC-based setpoint control yields cost savings closely aligned with energy savings (around 9%).

4.9. Tuning Facility Parameters

A separate tuning experiment samples candidate facility and ambient parameters (e.g., setpoint, cooling fraction, COP bounds, ambient baseline and swing, economizer gain) and evaluates average energy reductions across multiple seeds. The best candidate in our sweep yields a mean reduction of approximately 9.41%, slightly

improving upon the default configuration without introducing instability. Such parameter sweeps can be automated for different climates or hardware generations.

4.10. Summary of Findings

Our results show that even a relatively simple combination of carbon-aware scheduling and MPC-based setpoint control can deliver meaningful energy and emissions reductions in a simulated GQDC:

- Facility energy reductions of about 9% are achieved with tight confidence intervals and strong bootstrap evidence against the null hypothesis of no improvement.
- Carbon-aware deferrals can reduce emissions by 4–13% depending on aggressiveness, with operational overheads quantified as deferrals per job.
- Nonlinear COP modeling modestly enhances savings and increases realism, highlighting the role of accurate facility models.
- Fairness analysis reveals how deferrals can disproportionately impact specific job classes, underscoring the need for explicit fairness objectives.
- Robustness checks under high-load and heat-wave scenarios suggest that the qualitative benefits of MPC and carbon-aware scheduling persist under stress.

4.11. Relation to Prior Work

Compared to industrial carbon-aware systems for classical data centers [6, 9, 10], our framework operates at smaller scale and relies on synthetic signals, but it highlights similar design tensions between temporal shifting, SLA preservation, and fairness. Relative to prior work on energy efficiency and cooling optimization [11-13], our contribution is to integrate a simplified yet tunable facility model into a workload-level scheduling pipeline for quantum workloads.

Our earlier work on deep learning for solar power forecasting [5] focused on improving short-term solar-power forecasts via LSTM models trained on PVGIS–SARAH3 data [1-4]. The present study can be viewed as a complementary effort: given a carbon signal (forecasted or synthetic), how should a quantum data center adjust scheduling and cooling to exploit that information? Future work could integrate explicit forecast models and quantify the impact of forecast errors on scheduling decisions.

4.12. Limitations and Threats to Validity

Several limitations of our study warrant caution:

- **Synthetic signals and workloads.** We use synthetic carbon, price, ambient, and workload traces. While these are parameterized to mimic realistic diurnal patterns, they cannot capture all intricacies of real grids or quantum users.
- **Reduced-order facility model.** Our COP models and load aggregation are simplified representations of complex cooling and cryogenic systems. Detailed CFD or plant-level simulations may reveal additional interactions.
- **Single-site setting.** We consider a single GQDC; multi-site scheduling with geospatial carbon variation introduces additional opportunities and constraints [10, 6].
- **Limited policy space.** We focus on simple deferral heuristics, FIFO/EDF schedulers, and a lightweight MPC. More advanced optimization (e.g., MILP, reinforcement learning, or policy search) could uncover better trade-offs [13, 7].

Despite these limitations, the combination of ablations (e.g., nonlinear COP), stress scenarios, and bootstrap testing supports the qualitative robustness of the main findings.

4.13. Future Work

Future directions include:

- Integrating data-driven carbon and price forecasts and quantifying sensitivity to forecast errors.
- Extending the scheduler to handle precedence constraints in multi-stage quantum workflows, drawing on recent work in carbon- and precedence-aware scheduling [8, 9].
- Incorporating explicit fairness objectives (e.g., constraints on Jain index or Gini coefficient) into the optimization.
- Scaling to multi-site GQDCs with latency and bandwidth constraints, enabling joint spatial and temporal carbon-aware orchestration.
- Validating the reduced-order facility model against detailed simulations or measurements from prototype quantum data centers.

5. Conclusion

We introduced a simulation-based framework for green quantum data centers that couples carbon-aware job deferrals with model predictive cooling setpoint control. Using synthetic but tunable workloads and signals, we quantified facility energy savings, emissions reductions, queueing behavior, fairness across job classes, scheduler baselines, stress scenarios, and economic impacts.

Our experiments show that MPC can reduce facility energy by roughly 9% with tight confidence intervals and strong bootstrap evidence, while carbon-aware deferrals yield additional emissions reductions in the 4–13% range depending on aggressiveness. Nonlinear COP modeling and stress tests further demonstrate the importance of facility realism and robustness analysis. The full artifact—including code, configuration, figures, tables, and document builders—enables one-click reproduction and extension, offering a reusable template for future work on sustainable quantum computing infrastructure.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

No external funding was received for this research.

Acknowledgments

The authors acknowledge that this work received no external funding and used no proprietary datasets. The simulation framework is informed by general concepts from the solar forecasting literature and carbon-aware data center research; however, all workloads, signals, and experiments in this study are synthetic and designed to be publicly reproducible.

Availability of Data and Materials

An artifact containing the full simulation framework, configuration files, experiment drivers, and scripts to regenerate all figures and tables is available at:

https://github.com/ahmadrezadehghan/Carbon-Aware-Machine-Learning-for-Energy-Efficient-Quantum-Data-Centers.git

All simulation, scheduling, control, and plotting code required to reproduce the experiments and figures is provided as an artifact-ready implementation (modules, experiment drivers, and figure scripts).

Author Contributions

Sahand Heidary contributed to the design of the carbon-aware deferral policies, analysis scripts, and evaluation workflow, and assisted with figure preparation. Seyed Ahmadreza Dehghanian conceived and designed the study, implemented the GQDC simulation framework, conducted the experiments, and drafted the manuscript. Reza Sari Aslani contributed to the system modeling assumptions and helped refine the methodological framing and presentation of results. Rahim Zahedi supervised the research, provided technical guidance on energy systems and carbon-aware operation, and critically revised the manuscript. All authors reviewed and approved the final version.

References

[1] Krinner S, Storz S, Kurpiers P, Magnard P, Heinsoo J, Keller R, *et al*. Engineering cryogenic setups for 100-qubit scale superconducting circuit systems. EPJ Quantum Technol. 2019; 6(1): 2. https://doi.org/10.1140/epjqt/s40507-019-0072-0

[2] Deng X, Schulz M, Schulz L. Power consumption and energy efficiency of quantum computing platforms in high performance computing integration. In: International Conference on Parallel Processing and Applied Mathematics; 2024 Sep 9; Cham. Cham: Springer Nature Switzerland; 2024. p. 325-37. https://doi.org/10.1007/978-3-031-85700-3_24

[3] Zou Y, Keskin B, Taylor GG, Li Z, Wang J, Alarcon E, *et al*. Power delivery for cryogenic scalable quantum applications: challenges and opportunities. arXiv:2511.13965 [cs]; 2025 Nov 17.

[4] Xu Y, Tian M, Yan C, Li D, Deng T, Guo P, *et al*. A data center energy efficiency optimization method based on optimal temperature control of designated active servers. Energy Build. 2025; 345: 116126. https://doi.org/10.1016/j.enbuild.2025.116126

[5] Safari A, Sorouri H, Rahimi A, Oshnoei A. A systematic review of energy efficiency metrics for optimizing cloud data center operations and management. Electronics. 2025; 14(11): 2214. https://doi.org/10.3390/electronics14112214

[6] Li X, Zhang Z, Chen Y, He R. A review on AI-driven optimization of data center energy efficiency and thermal management. Int J Appl Sci. 2025; 8(3): 108. https://doi.org/10.30560/ijas.v8n3p108

[7] Radovanović A, Koningstein R, Schneider I, Chen B, Duarte A, Roy B, *et al*. Carbon-aware computing for datacenters. IEEE Trans Power Syst. 2023; 38(2): 1270-80. https://doi.org/10.1109/TPWRS.2022.3173250

[8] Bostandoost R, Hanafy WA, Lechowicz A, Bashir N, Shenoy P, Hajiesmaili M. Data-driven algorithm selection for carbon-aware scheduling. ACM SIGENERGY Energy Inform Rev. 2024; 4(5): 148-53. https://doi.org/10.1145/3727200.3727222

[9] Liu W, Yan Y, Sun Y, Mao H, Cheng M, Wang P, *et al*. Online job scheduling scheme for low-carbon data center operation: an information and energy nexus perspective. Appl Energy. 2023; 338: 120918. https://doi.org/10.1016/j.apenergy.2023.120918

[10] Acun B, Lee B, Kazhamiaka F, Maeng K, Gupta U, Chakkaravarthy M, *et al*. Carbon explorer: a holistic framework for designing carbon-aware datacenters. In: Proc 28th ACM Int Conf Archit Support Program Lang Oper Syst (ASPLOS); 2023 Jan 27. p. 118-32. https://doi.org/10.1145/3575693.3575754

[11] Wiesner P, Behnke I, Kilian P, Steinke M, Kao O. Vessim: a testbed for carbon-aware applications and systems. ACM SIGENERGY Energy Inform Rev. 2024; 4(5): 59-66. https://doi.org/10.1145/3727200.3727210

[12] Maji D, Pfaff B, PR V, Sreenivasan R, Firoiu V, Iyer S, *et al*. Bringing carbon awareness to multi-cloud application delivery. In: Proc 2nd Workshop Sustainable Computer Systems; 2023 Jul 9. p. 1-6. https://doi.org/10.1145/3604930.3605711

[13] Takci MT, Qadrdan M, Summers J, Gustafsson J. Data centres as a source of flexibility for power systems. Energy Rep. 2025; 13: 3661-71. https://doi.org/10.1016/j.egyr.2025.03.020

[14] Jain RK, Chiu DM, Hawe WR. A quantitative measure of fairness and discrimination. Hudson (MA): Digital Equipment Corporation, Eastern Research Laboratory; 1984 Sep 26.

[15] Ware R, Mukerjee MK, Seshan S, Sherry J. Beyond Jain's fairness index: setting the bar for the deployment of congestion control algorithms. In: Proc 18th ACM Workshop Hot Topics in Networks; 2019 Nov 14. p. 17-24. https://doi.org/10.1145/3365609.3365855

[16] Zapletal A, Kuipers F. Slowdown as a metric for congestion control fairness. In: Proc 22nd ACM Workshop Hot Topics in Networks; 2023 Nov 28. p. 205-12. https://doi.org/10.1145/3626111.3628185

[17] Dehghanian SA, Heidary S, Shams D, Pour AM, Zahedi R. Deep learning for solar power forecasting: a robust stacked LSTM algorithm for operational applications. Edelweiss Appl Sci Technol. 2025; 9(10): 1591-600. https://doi.org/10.55214/2576-8484.v9i10.10722

[18] Shehabi A, Smith S, Sartor D, Brown R, Herrlin M, Koomey J, *et al*. United States data center energy usage report. Berkeley (CA): Lawrence Berkeley National Laboratory; 2016.

[19] Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy-use estimates. Science. 2020 Feb; 367(6481): 984-6. https://doi.org/10.1126/science.aba3758

[20] Zhang J, Hodge BM, Florita A, Lu S, Hamann HF, Banunarayanan V. Metrics for evaluating the accuracy of solar power forecasting. Golden (CO): National Renewable Energy Laboratory; 2013 Oct 1.

[21] Yang D, Kleissl J. Solar irradiance and photovoltaic power forecasting. Boca Raton (FL): CRC Press; 2024 Feb 5. https://doi.org/10.1201/9781003203971

[22] Xie Y, Habte A, Sengupta M, Vignola F. An evaluation of the spectral irradiance data from the NSRDB. Golden (CO): National Renewable Energy Laboratory; NREL/TP-5D00-80439. Available from: https://www.nrel.gov/docs/fy21osti/80439.pdf

[23] Tongsopit S, Junlakarn S, Chaianong A, Overland I, Vakulchuk R. Prosumer solar power and energy storage forecasting in countries with limited data: the case of Thailand. Heliyon. 2024; 10(2): e23997. https://doi.org/10.1016/j.heliyon.2024.e23997

[24] Fleschutz M, Murphy MD. elmada: dynamic electricity carbon emission factors and prices for Europe. J Open Source Softw. 2021; 6(66): 3625. https://doi.org/10.21105/joss.03625

[25] Hawkes AD. Estimating marginal CO_2 emissions rates for national electricity systems. Energy Policy. 2010; 38(10): 5977-87. https://doi.org/10.1016/j.enpol.2010.05.053

[26] Tranberg B, Corradi O, Lajoie B, Gibon T, Staffell I, Andresen GB. Real-time carbon accounting method for the European electricity markets. Energy Strategy Rev. 2019; 26: 100367. https://doi.org/10.1016/j.esr.2019.100367

[27] Zhang S, Zhao W, Zhu B, Yan C, Song X, Jiang H, *et al*. A near-real-time daily European power consumption and carbon intensity dataset (ECON-PowerCI). Sci Data. 2025; 12(1): 1693. https://doi.org/10.1038/s41597-025-05978-7

[28] Scholta HF, Blaschke MJ. Temporal matching as an accounting principle for green electricity claims. Nat Commun. 2025; 16(1): 9280. https://doi.org/10.1038/s41467-025-65125-z

[29] Lazic N, Boutilier C, Lu T, Wong E, Roy B, Ryu MK, *et al*. Data center cooling using model-predictive control. In: Advances in Neural Information Processing Systems. Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018); 2018 Dec 3–8; Montréal, Canada. Red Hook (NY): Curran Associates; 2018.

[30] Goiri Í, Nguyen TD, Bianchini R. Coolair: temperature- and variation-aware management for free-cooled datacenters. ACM SIGPLAN Not. 2015; 50(4): 253-65. https://doi.org/10.1145/2775054.2694378