sustainability

MDPI

Article

Sustainable Grid Integration of AI Data Centers: Carbon-Aware Co-Optimization of Training Workloads and Battery Storage Under Renewable Uncertainty

Junqing Zhang [1,*] and Yi Wang [2]

[1] Efrei Research Lab, Efrei Paris Panthéon Assas University, 30–32 Av. de la République, 94800 Villejuif, France
[2] School of Electrical and Automation Engineering, Faculty of Engineering, Wuhan University, Wuhan 430072, China
* Correspondence: junqing.zhang@efrei.net

Abstract

The electricity demand of artificial intelligence (AI) data centers is growing faster than the grids hosting them can decarbonize, yet carbon-aware computing is rarely audited at the system level: workloads follow scalar-price or carbon signals, whose claimed savings need not survive network-wide dispatch. We develop a carbon-aware co-optimization framework that schedules checkpoint intervals, workload migration, dynamic voltage and frequency scaling, and behind-the-meter storage jointly with unit commitment, pricing system CO_2 emissions beyond the market-internalized level, with split-conformal scenario bands and an out-of-sample cost certificate. On IEEE 39/118-bus systems mapped to 2025 French and German measurements, carbon-priced dispatch reduces emissions by 29.5% at a generation-cost premium below 0.01% (the carbon-signal headroom of this mapped system); system-level decomposition attributes more than 99% of this reduction at reference penetration to generation redispatch, an operator-side instrument. AI-load flexibility adds 0.23% of system emissions today, equivalent to about one quarter of the data-center fleet's own carbon footprint, and 3.3% at gigawatt scale, where it cuts worst-day load shedding by up to 49%. The endogenous checkpoint policy recovers the reliability-optimal interval, whereas following the average carbon-intensity signal can increase system-level emissions. Conformal calibration maintains at-or-above-nominal coverage (97.5%); empirical bands under-cover (86.0%).

Keywords: data centers; artificial intelligence workloads; carbon-aware scheduling; battery energy storage; renewable energy integration; two-stage stochastic optimization

Check for updates

Academic Editor: Zhiyuan Jiang

Received: 31 August 2026
Revised: 11 September 2026
Accepted: 14 September 2026
Published: 16 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Data centers have become one of the fastest-growing electricity loads worldwide. The International Energy Agency estimates that data centers, cryptocurrencies, and artificial intelligence (AI) together consumed roughly 460 TWh in 2022 and this could more than double to over 1000 TWh by 2026, with AI as the principal growth driver [1,2]. This growth is notable because it reverses a decade-long trend in which efficiency gains kept data-center electricity use nearly flat despite rapidly growing demand for digital services [3,4]. A single large-model training run can emit hundreds of tonnes of CO_2 [5], and the research community has responded with calls for "Green AI": reporting and reducing the carbon cost of computation [6–8]. Hyperscale operators have pledged around-the-clock carbon-free

operation [9], yet the pace of AI capacity build-out is beginning to strain the physical limits of the power systems that host it.

From the power-system perspective, AI data centers are not merely large loads, they are concentrated, fast-growing, and spatially mobile, and their grid integration has become a sustainability challenge in its own right, touching Sustainable Development Goals (SDGs) 7 (affordable and clean energy) and SDG 13 (climate action) simultaneously. This journal has documented the building blocks of sustainable data-center operation: energy-mix and circular-energy strategies [10]; green data-center frameworks for energy optimization and carbon-footprint reduction [11]; critical success factors for green-energy integration [12]; the AI energy-hunger paradox [13]; and smart-grid-connected data-center energy management [14], including multi-energy-grid integration validated in operational environments [15]. Beyond data centers, numerical simulation and optimization models underpin design and operation across the wider energy sector, from multivariate coupling models of enhanced geothermal reservoirs [16] and fracture-propagation simulations of oil-shale reservoirs [17] to temperature–pressure optimization of gas-hydrate production [18]; the present work brings the same model-driven approach to the grid side of AI computing. A common thread of this literature is that the data center is treated as a passive consumer whose supply side must be greened; the complementary question, namely, how the grid should dispatch around an AI load that can itself move in space, shift in time, and reshape its own power profile, remains open.

That such flexibility exists has been established by two research streams. In the Internet-systems literature, geographical load balancing exploits price and renewables diversity across regions [19–21], dynamic right-sizing matches server provisioning to the load [22], and data centers have been shown capable of offering demand response [23–25] and ancillary services [26–28]. In the carbon-aware-computing literature, hyperscalers shift flexible workloads toward hours and locations with cleaner electricity [29,30], and carbon-aware inference services trade model quality against emissions [31]. Behind-the-meter battery storage adds a third flexibility channel, whose value stacking for renewable integration and peak management is well documented [32–36]. These streams, however, stop at the data-center boundary: the power system is represented, if at all, as a scalar price or carbon signal, and the workload is represented as an exogenous, freely shapeable energy quantity.

What these streams leave open is the grid side of the problem. The power shape of AI training is itself decision-dependent, a dependence absent from grid models. Synchronous large-model training alternates between compute phases and checkpoint events whose depth, duration, and frequency are set by a reliability policy [37–39]; production reports from Llama-3-class and MegaScale-class clusters show that failure rates, checkpoint overheads, and preemption cascades materially shape facility power at production scale [40,41]. The checkpoint interval is conventionally chosen to minimize wasted compute alone, ignoring the fact that denser checkpointing raises facility power exactly when the grid can least afford it. Carbon-aware scheduling is meanwhile network-agnostic: shifting load toward a low-carbon region changes redispatch, flows, and marginal emissions in ways a scalar signal cannot capture, so claimed reductions need not materialize at the system level. Uncertainty treatment presents a parallel picture. Robust and adaptive unit commitment under wind uncertainty is mature [42–45], chance-constrained formulations address probabilistic bounds [46], and data-driven distributionally robust counterparts model ambiguity directly [47–51]. Conformal prediction adds distribution-free, finite-sample coverage guarantees that have begun to enter renewable-energy forecasting [52–56]; yet scenario sets for data-center-coupled dispatch are still generated without such calibration, leaving the out-of-sample reliability of the resulting schedules unverified.

We close these gaps with a carbon-aware co-optimization framework for the sustainable grid integration of AI training data centers. The framework formulates a two-stage stochastic unit-commitment (UC) problem that jointly coordinates an endogenous checkpoint-shaped training load, geographical workload migration, DVFS-based temporal shaping, and behind-the-meter battery storage, while explicitly accounting for system-level CO_2 emissions and generation costs. The checkpoint interval enters as a day-ahead grid-side decision, coupling AI training reliability with power-system scheduling. Uncertainties are characterized through split conformal prediction over bootstrapped historical forecast errors, yielding an explicit out-of-sample performance certificate. Evaluation uses measured 2025 French and German zonal data mapped onto the IEEE 39- and 118-bus systems.

The main contributions are fourfold:

- We develop a grid-coupled dispatch model in which the checkpoint interval is treated as an endogenous scheduling decision, jointly optimized with workload migration, DVFS, and battery operation, and we use it to decompose, at system level, where the emission benefit of carbon-aware dispatch actually comes from: at current penetration, more than 99% is delivered by carbon-priced generation redispatch, an operator-side instrument available irrespective of data-center cooperation. The decomposition identifies the generation side, not the load side, as the dominant near-term decarbonization lever for grids hosting AI loads.
- We show that system-level shares understate the value of AI-load flexibility. Expressed against the data-center fleet's own consumption-weighted carbon footprint, the flexibility channels abate about one quarter of it, a proportion that is roughly constant across scales, while their system-level share grows super-linearly from 0.02% to 3.3% of emissions as penetration rises; at gigawatt scale they also become an adequacy resource, cutting worst-day load shedding by up to 49%. The endogenous checkpoint policy recovers the Daly–Young reliability optimum throughout, at no reliability-cost premium.
- We formulate a risk-averse two-stage stochastic optimization framework in which split conformal prediction calibrates uncertainty scenarios from historical forecast errors. This provides an explicit out-of-sample cost certificate, whose empirical coverage is further validated using a full year of held-out data.
- In a fully reproducible European case study on measured 2025 French and German data, we show that workload scheduling based solely on exogenous carbon-intensity signals, the prevailing carbon-aware practice, can increase system-level emissions, whereas embedding carbon emissions directly in system dispatch lets workload decisions account for their network-wide generation consequences and yields reliable carbon reduction (Section 4.1).

The rest of the paper proceeds from model to evidence: Section 2 develops the system model and data, Section 3 the experimental design, Section 4 the results, and Sections 5 and 6 the discussion and conclusions.

2. Materials and Methods

2.1. System Overview and Modeling Assumptions

We consider a transmission-level power system hosting a set $\mathcal{M}$ of geographically distributed AI training data centers (DCs), each equipped with a behind-the-meter battery energy storage system (BESS). The system operator schedules conventional generation, wind acceptance, and inter-DC workload allocation over a day-ahead horizon $\mathcal{T} = \{1, \ldots, 24\}$ with $\Delta t = 1$ h, under uncertainty in wind generation and electrical load. The setting departs from classical economic dispatch in the load itself, its flexibility, and its objective. The AI

training load has an endogenous, decision-dependent power shape driven by checkpointing; computing demand is spatially shiftable through workload migration and temporally shiftable through dynamic voltage and frequency scaling (DVFS); and the environmental objective is explicit: grid CO_2 emissions enter the objective alongside cost, and the marginal value of flexibility is evaluated against both.

We represent the physical network by a benchmark DC power-flow topology (IEEE 39- and 118-bus systems, used as the published MATPOWER case files in Case Format 2 [57]) and map two electricity-market zones, France (FR) and Germany (DE), onto it as a reduced-form representation of the coupled European system. The mapping follows three documented conventions. The generator set $\mathcal{G}$ is partitioned into an FR fleet and a DE fleet by merit order: the cheapest units, up to a 40% capacity share, form the low-carbon FR fleet, reflecting the nuclear-dominated French mix, and the remainder form the DE fleet. Zone membership then propagates from generators to buses by breadth-first expansion from the FR fleet buses, so that nodal loads follow the zonal load profile of 2025 measurements. Each unit's time-varying marginal cost and emission rate are anchored to the common coupled-market day-ahead price and to the zonal average carbon intensity, respectively (Section 2.5); the price is scaled by a static per-unit merit factor (the MATPOWER linear cost coefficient normalized to unit mean within each zone), so that the benchmark merit order differentiates units without introducing an artificial cross-zone cost bias. Wind farms are placed at the highest-load buses of each zone with an aggregate capacity of 25% of peak load, and are driven by measured 2025 zonal wind profiles. This reduced-form design follows the common practice of driving benchmark networks with real market and resource data; it preserves network physics while making the carbon and price signals fully empirical.

Six working assumptions are adopted:

- (A1) GPU failures arrive as a Poisson process with per-GPU rate $\lambda_m(N_m)$, the standard checkpointing-model idealization [37,38], with the rate calibrated to public production records (Section 3).
- (A2) The checkpoint overhead is small relative to the checkpoint interval, which legitimizes the average-power representation of the checkpoint dip.
- (A3) Server counts are treated as continuous variables, a relaxation whose rounding error is at most one server per DVFS state, at most 0.1% of N_m at the smallest tested scale.
- (A4) The BESS operates energy-neutrally over each day with constant round-trip efficiency.
- (A5) Day-ahead prices are known at scheduling time.
- (A6) The network physics is represented by DC power flow.

Each assumption is revisited where it is used.

2.2. Checkpoint-Shaped AI Training Load

Each DC $m \in \mathcal{M}$ hosts synchronous distributed training on N_m GPUs. The training state is persisted periodically: every $f_m\tau$ seconds, where f_m is the checkpoint interval in iterations and τ the per-iteration time, all GPUs stall for a dip of relative depth ρ_{dip} and duration ω. At the hourly dispatch resolution, the checkpoint dips enter as an average power reduction and, more importantly for grid operation, as an overhead that inflates the compute that must be served; the wasted GPU-hours per day are decomposed into stall and recomputation terms:

$$\mathcal{W}_m(f_m) = \underbrace{\lambda_m(N_m)\, N_m^2\, \frac{24 f_m \tau}{7200}\,(1+\gamma_{\text{cascade}})}_{\text{recomputation after failures}} + \underbrace{N_m\, \frac{24\omega}{f_m\tau}}_{\text{checkpoint stalls}} \tag{1}$$

where $\lambda_m(N_m)$ is the per-GPU failure rate; and γ_{cascade} accounts for preemption cascades, in which a failed large job evicts smaller ones. In the recomputation term, $24 f_m \tau$ is the checkpoint-interval exposure in seconds per GPU-day and the divisor $7200 = 2 \times 3600$ converts it to GPU-hours while carrying the factor $\frac{1}{2}$ of the uniform failure-position assumption: each failure wastes on average half an interval of recomputed work [37,38]. Both terms are monetized symmetrically at the GPU-hour rental price c^{gpu}, giving the training-reliability cost $C_m^{\text{rel}}(f_m) = c^{\text{gpu}} \mathcal{W}_m(f_m)$. Symmetric accounting preserves the classical trade-off: minimizing Equation (1) yields the optimal interval [37,38]

$$f_m^{\text{DY}} = \max\left(1,\ \tfrac{1}{\tau}\sqrt{7200\,\omega / \left(\lambda_m(N_m) N_m (1 + \gamma_{\text{cascade}})\right)}\right) \tag{2}$$

which reduces to the classical Daly–Young formula when $\gamma_{\text{cascade}} = 0$; at the calibrated $\gamma_{\text{cascade}} = 0.16$, the exact minimizer shifts by the factor $1/\sqrt{1.16} \approx 0.93$, which leaves its nearest candidate in $\mathcal{F}$ unchanged at every scale we test. We use Equation (2) both as a baseline policy (Section 3) and as an analytic verification anchor for the implementation (recent evidence qualifies when Young/Daly-type intervals remain appropriate [58]). In this work f_m is not fixed at f_m^{DY} but is a day-ahead grid-side decision: denser checkpointing (smaller f_m) raises stall energy and thus DC power, while sparser checkpointing raises recomputation exposure; the dispatchable trade-off between the two affects the grid price reliability against carbon.

The checkpoint interval is selected from a candidate set $\mathcal{F}$ by binaries $y_{m,k}$ with $\sum_{k \in \mathcal{F}} y_{m,k} = 1$, so that $f_m = \sum_k k\, y_{m,k}$ and the overhead fraction

$$\eta_m^{\text{ov}} = \sum_{k \in \mathcal{F}} \frac{\rho_{\text{dip}}\,\omega}{k\,\tau}\, y_{m,k} \tag{3}$$

is linear in the decision. The reliability cost (1) is likewise linearized exactly through the same binaries, since $\mathcal{W}_m$ depends on f_m only through the values in $\mathcal{F}$.

2.3. Data-Center Migration, DVFS, and Service Quality

2.3.1. Workload Migration

Workload can be shifted between DCs over a set $\mathcal{A}$ of migration arcs with a one-period communication delay. With $\phi_{ij,t}$ the signed workload transferred on arc (i, j) at period t (positive into i), the served workload of DC m is

$$W_{m,t} = W_{m,t}^0 + \sum_{j:(m,j) \in \mathcal{A}} \phi_{mj,t-1}, \qquad \phi_{ij,t} + \phi_{ji,t} = 0 \tag{4}$$

$$-\overline{\phi}_{ij}\,\sigma_{ij,t}^- \le \phi_{ij,t} \le \overline{\phi}_{ij}\,\sigma_{ij,t}^+, \qquad \sigma_{ij,t}^+ + \sigma_{ij,t}^- \le 1 \tag{5}$$

where $W_{m,t}^0$ is the original workload demand, $\overline{\phi}_{ij}$ the arc capacity, and $\sigma^+, \sigma^- \in \{0, 1\}$ the direction selectors. Migration is not free: a transfer cost c^{mig} per GPU-hour migrated, entering the objective as $c^{\text{mig}} \sum_{(i,j),t} |\phi_{ij,t}|$, prices wide-area data movement and gives the spatial reallocation a break-even threshold against the carbon savings it enables. Migration relocates electricity consumption between market zones while each DC–grid exchange remains an ordinary load at its new location. This property makes carbon-aware migration a meaningful instrument under the wide zonal carbon contrast between FR and DE documented in Section 2.5.

2.3.2. Server States and Computing Power

Servers of DC m operate in DVFS states $h \in \mathcal{H}$ with per-server power factor κ_h^P and service-rate factor κ_h^S relative to the full rate. With $n_{m,h,t}$ servers in state h,

$$\sum_{h\in\mathcal{H}} n_{m,h,t} \le N_m, \qquad P_{m,t}^{\mathrm{IT}} = p^{\mathrm{gpu}} \sum_{h\in\mathcal{H}} \kappa_h^P \, n_{m,h,t} \tag{6}$$

where p^{gpu} is the measured per-GPU steady power. The total DC grid draw combines IT power, cooling through a static power usage effectiveness (PUE) PUE_m, and the behind-the-meter BESS exchange:

$$P_{m,t}^{\mathrm{dc}} = \mathrm{PUE}_m \, P_{m,t}^{\mathrm{IT}} - P_{m,t}^{\mathrm{dis}} + P_{m,t}^{\mathrm{ch}} \tag{7}$$

2.3.3. Service Quality

Computing capacity must cover the served workload inflated by the checkpoint overhead, with a service headroom that approximates an M/M/1-type mean-response-time constraint:

$$\sum_{h\in\mathcal{H}} \kappa_h^S \, n_{m,h,t} \;\ge\; (1+\epsilon)\, W_{m,t} + \eta_m^{\mathrm{ov}} \, W_{m,t}^0 \tag{8}$$

where ϵ is the quality-of-service (QoS) headroom keeping the mean response time within the admissible delay. The overhead term is applied to the baseline workload W^0 (a constant), keeping the constraint linear; the approximation error is bounded by the overhead fraction (at most 5.5% over $\mathcal{F}$) times the migration share (at most 0.3), hence below 2% of the workload. The term is evaluated at the endogenous checkpoint choice, whereas baselines with an exogenously fixed interval omit it; this conservative asymmetry is revisited in Section 4.6.

2.4. Behind-the-Meter Battery Storage

Each DC operates a BESS with power rating P_m^{bess}, energy rating $E_m^{\mathrm{cap}} = \bar{h}\, P_m^{\mathrm{bess}}$ ($\bar{h} = 2$ h), single-trip efficiencies $\eta_{\mathrm{c}} = \eta_{\mathrm{d}}$, state-of-charge window $[\underline{\mathrm{soc}}, \overline{\mathrm{soc}}]$, and daily energy neutrality:

$$\mathrm{SOC}_{m,t+1} = \mathrm{SOC}_{m,t} + \frac{\Delta t}{E_m^{\mathrm{cap}}}\left(\eta_{\mathrm{c}} P_{m,t}^{\mathrm{ch}} - \frac{P_{m,t}^{\mathrm{dis}}}{\eta_{\mathrm{d}}}\right), \quad \mathrm{SOC}_{m,1} = \mathrm{SOC}_{m,25} = 0.5 \tag{9}$$

Nameplate capacity follows the checkpoint-driven sizing rule $P_m^{\mathrm{bess}} = 0.38$ kW per GPU, motivated by dip-masking requirements, with a sizing multiplier explored in sensitivity analysis. Because $\eta_{\mathrm{c}}\eta_{\mathrm{d}} < 1$ and throughput is priced by a degradation term c^{deg}, simultaneous charging and discharging is never optimal and no complementarity binaries are needed [59].

2.5. Zonal Carbon and Price Signals from Public Data

All exogenous time series are hourly 2025 measurements from public sources that require no registration or API keys, following the frictionless open-data practice of [60]. For France, the RTE éCO2mix open dataset [61] provides the official hourly average carbon intensity $I_{\mathrm{FR}}(t)$ ($\mathrm{gCO_2/kWh}$), zonal load, wind generation, and the day-ahead load forecast used for forecast-error statistics. For Germany, the SMARD platform of the Bundesnetzagentur [62] provides the day-ahead price $\pi(t)$, total load, and per-fuel genera-

tion; the hourly average carbon intensity is computed as the generation-weighted mean of operational emission factors consistent with IPCC/EEA inventories [63,64]:

$$I_{\mathrm{DE}}(t) = \frac{\sum_{f \in \mathcal{F}^{\mathrm{fuel}}} q_f(t)\, e_f}{\sum_{f \in \mathcal{F}^{\mathrm{fuel}}} q_f(t)} \tag{10}$$

with $q_f(t)$ the measured per-fuel generation and e_f the emission factors of Table 1. Minor contributors below 2% of annual energy (hydro, pumped storage, and miscellaneous categories) are omitted, which overstates I_{DE} by at most the omitted energy share (below 2% of annual generation). Day-ahead prices for both zones are represented by the DE-LU auction price, reflecting the strong price coupling of the Single Day-Ahead Coupling; because a common price removes spatial arbitrage, the spatial carbon instrument is exercised against the migration transfer cost, whose sensitivity is reported in Section 4. The 2025 contrast between the two zones is the empirical basis of the carbon-aware instrument: Figure 1 shows the FR series averaging 17 gCO_2/kWh, while the computed DE series averages 366 gCO_2/kWh, twenty-one times higher, with a pronounced midday solar dip.

Table 1. Operational emission factors used for the DE carbon-intensity computation [63,64].

Fuel	e_f **[gCO$_2$/kWh]**
Lignite	1100
Hard coal	900
Natural gas	490
Biomass	230
Nuclear, hydro, wind, solar	0

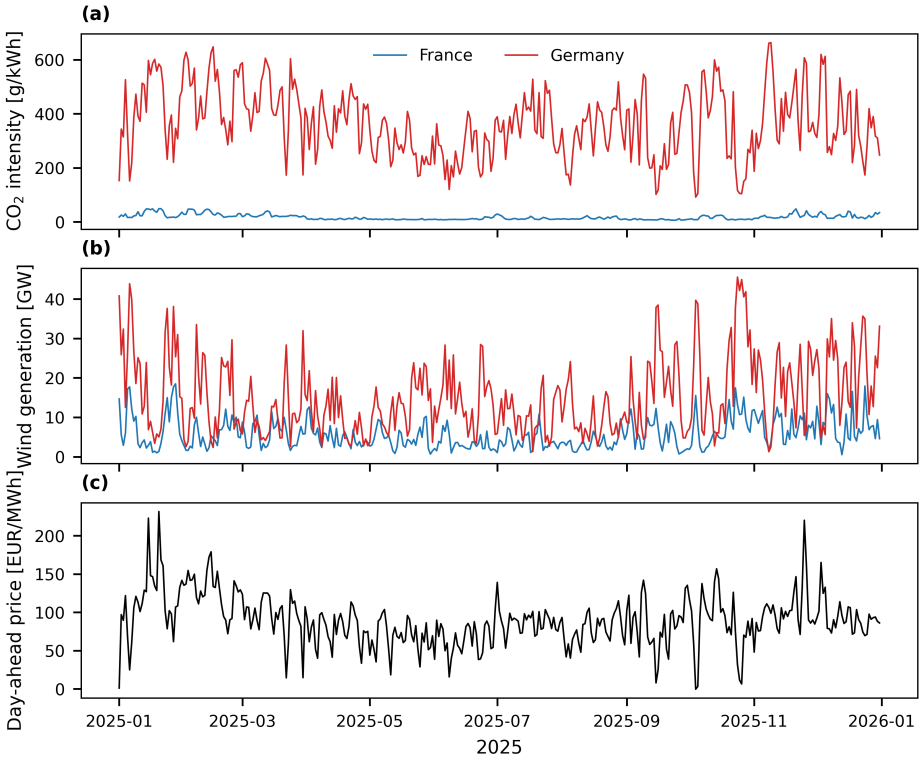


Figure 1. The 2025 zonal data overview: (**a**) daily-mean grid CO_2 intensity; (**b**) daily-mean wind generation; (**c**) daily-mean day-ahead price (DE-LU).

Scope of the Carbon Price

Observed day-ahead prices already internalize the cost of European Union Emissions Trading System (EU ETS) allowances surrendered by fossil generators; the 2025 average auction price was 73.43 EUR/tCO$_2$ [65]. The coefficient π_{CO_2} in Equation (15) is therefore an incremental shadow price on emissions beyond the market-internalized level, not a stand-alone carbon price, and the "carbon-blind" reference B1 ($\pi_{\mathrm{CO}_2} = 0$) corresponds to dispatch under prices as observed, not to a world without carbon policy. The Pareto sweep of Section 4.3 thus traces the effect of strengthening the carbon signal above its market-embedded baseline; a sensitivity variant that strips the internalized allowance cost from the price anchor ($\pi_{\mathrm{net}}(t) = \pi(t) - p_{\mathrm{EUA}} I_{\mathrm{DE}}(t)/10^3$, $p_{\mathrm{EUA}} = 70$ EUR/t, a rounded value of the 73.43 EUR/tCO$_2$ average quoted above) separates the explicit carbon price from the embedded one; Section 4.3 reports the comparison.

2.6. Network and Generation Model

Conventional units are scheduled with commitment states $z_{g,t}$, start-up/shut-down indicators $u_{g,t}, v_{g,t}$, output limits, and ramp limits; network flows follow DC power flow with line capacities:

$$\underline{P}_g z_{g,t} \le P^g_{g,t} \le \overline{P}_g z_{g,t}, \quad u_{g,t} - v_{g,t} = z_{g,t} - z_{g,t-1}, \quad u_{g,t} + v_{g,t} \le 1 \tag{11}$$

$$P^{\mathrm{line}}_{\ell,t} = \frac{\theta_{b(\ell),t} - \theta_{b'(\ell),t}}{X_\ell}, \qquad |P^{\mathrm{line}}_{\ell,t}| \le \overline{F}_\ell \tag{12}$$

$$\sum_{g \in \mathcal{G}_b} P^g_{g,t} + \sum_{w \in \mathcal{W}_b} P^w_{w,t} + P^{\mathrm{in}}_{b,t} + P^{\mathrm{shed}}_{b,t} = P^d_{b,t} + \sum_{m \in \mathcal{M}_b} P^{\mathrm{dc}}_{m,t} \tag{13}$$

where wind acceptance $P^w_{w,t} \le \overline{P}^w_{w,t}$ (curtailment is the slack), P^{shed} is penalized load shedding, and $P^{\mathrm{in}}_{b,t}$ collects net line inflows from Equation (12).

2.7. Two-Stage Risk-Averse Formulation

We cast the dispatch problem as a two-stage stochastic program. First-stage (here-and-now) decisions are the commitment schedule, the checkpoint intervals, the migration plan, and the BESS schedule. Second-stage (wait-and-see) decisions, taken under each scenario $\omega \in \Omega$, are generation redispatch, wind acceptance, load shedding, and DVFS server-state adjustment. The objective reads

$$\min \ \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \Phi_\omega + \lambda_{\mathrm{risk}} \, \mathrm{CVaR}_\alpha(\Phi) + \sum_{m \in \mathcal{M}} C^{\mathrm{rel}}_m(f_m) + C^{\mathrm{deg}} \tag{14}$$

$$\Phi_\omega = \sum_{t,g} c_{g,t} P^g_{g,t,\omega} + \pi_{\mathrm{CO}_2} \sum_{t,g} e_{g,t} P^g_{g,t,\omega} + \rho^{\mathrm{w}} \sum_{t,w} \left(\overline{P}^w_{w,t,\omega} - P^w_{w,t,\omega} \right) + \rho^{\mathrm{s}} \sum_{t,b} P^{\mathrm{shed}}_{b,t,\omega} \tag{15}$$

where $c_{g,t}$ and $e_{g,t}$ are the zonal-data-anchored marginal cost and emission rate, π_{CO_2} a carbon price swept for the Pareto analysis, $\rho^{\mathrm{w}}, \rho^{\mathrm{s}}$ penalty prices, and CVaR_α the conditional value at risk (CVaR) at level $\alpha = 0.9$ in its linear Rockafellar–Uryasev form,

$$\mathrm{CVaR}_\alpha(\Phi) = \min_\zeta \ \zeta + \frac{1}{(1-\alpha)\,|\Omega|} \sum_{\omega \in \Omega} \max\left(\Phi_\omega - \zeta, \, 0\right) \tag{16}$$

which introduces only $|\Omega| + 1$ auxiliary variables and keeps the model a mixed-integer linear program (MILP). The CVaR-regularized sample-average formulation is motivated by Wasserstein distributionally robust optimization, in which tail-risk regularization of the empirical expectation bounds the worst-case expectation over a ball of distributions [47]; the out-of-sample validation in Section 4 verifies that the cost certificates hold.

Scenarios are generated by bootstrapping historical forecast errors, stratified by hour of day, from the 2025 residual libraries: true day-ahead forecast errors where the source publishes forecasts (FR load), and persistence errors otherwise. The perturbation bands are calibrated by split conformal prediction [52]: residuals are split chronologically, within each hour-of-day residual pool, into calibration and verification halves, so that the earlier half of the record sets the conformal quantile with the finite-sample correction and the later half is held out and never used for tuning, giving distribution-free marginal coverage of the realized deviations. The ablation in Section 4 compares conformal-calibrated bands against empirical 90% quantiles.

3. Experimental Design, Baselines, and Metrics

We organize the evaluation into five experiments: E1 compares the proposed method against the eight baselines over the twelve representative days (Section 4.1); E2 sweeps the incremental carbon price (Section 4.3); E3 ablates the calibration and modeling choices on a single day (Section 4.4); E4 validates the cost certificate out of sample (Section 4.5); and E5 analyzes scale effects (Section 4.6).

3.1. Baselines

Eight baselines isolate the value of each mechanism: B1 *carbon-blind* (all flexibility on, $\pi_{CO_2} = 0$); B2 *no flexibility* (fixed workload, no DVFS or BESS, $f = f^{DY}$); B3 *BESS only*; B4 *migration+DVFS only*; B5 *Daly–Young checkpoint* (flexibility on, f fixed at Equation (2)); and B6 *signal following*, a literature-style carbon-chasing strategy in which a greedy migration plan shifts workload every hour toward the lowest average-intensity zone, up to arc capacity and the receiving site's headroom, with $f = f^{DY}$, no BESS, and a carbon-blind grid dispatch around it. Two further heuristics isolate the temporal and combined channels of signal following: B7 *temporal signal following*, which greedily shifts up to 20% of each zone's hourly workload from the highest- to the lowest-intensity hours of the day under day-long energy conservation, with no migration, no BESS, $f = f^{DY}$, and carbon-blind dispatch; and B8 *combined signal following*, which applies B7's time shift on top of B6's spatial migration plan. The proposed method co-optimizes all mechanisms with endogenous checkpoint intervals and the carbon-aware objective.

3.2. Metrics

We report system CO_2 (t/day), renewable curtailment (MWh/day), total cost (EUR/day), the training-reliability cost share, migration volume, BESS throughput, out-of-sample coverage of the cost certificate (evaluated separately on fresh same-day scenario draws and on held-out realized days, Section 4.5), and solver statistics.

3.3. Test Systems and DC Layouts

IEEE 39- and 118-bus networks; three DCs (one FR, two DE) at high-load buses; GPU scales $N \in \{10^3,\ 1.6 \times 10^4,\ 10^5,\ 3 \times 10^5\}$ per DC. Training-load parameters are anchored to public production records of large-scale training systems ($p^{gpu} = 598$ W, $\rho_{dip} = 48.5\%$, $\omega = 3.4$ s, $\tau = 15$ s, $\gamma_{cascade} = 0.16$) [40,41]; c^{gpu} is swept over reported GPU rental prices. The per-GPU failure rate is scale-independent, $\lambda_m(N) = 2.195 \times 10^{-5}$ failures per GPU-hour; each DC operates at $PUE_m = 1.3$, and its baseline workload is a 70% utilization of the installed GPUs, $W^0_{m,t} = 0.7\, N_m$ GPU-equivalents per hour.

3.4. Computational Setup

Twelve representative days of 2025, one per month, form the evaluation matrix: for each month we select the day whose daily-mean German wind generation is closest to

that month's median, so that the matrix spans all seasons without extreme-day selection (the selected dates are 25 January, 18 February, 25 March, 12 April, 3 May, 14 June, 2 July, 27 August, 7 September, 10 October, 22 November, and 14 December 2025). All experiments use $S = |\Omega| = 20$ sample-average-approximation (SAA) scenarios per dispatch day on the 39-bus system ($S = 10$ on the 118-bus system for tractability), with $\alpha = 0.9$ and $\lambda_{\text{risk}} = 0.5$ unless ablated; the remaining model and computational parameters are collected in Table 2. The MILP is solved with Gurobi 13.0.2 (academic license), called directly through its Python API from an implementation in Python 3.13.6 using NumPy 2.2.6, pandas 2.3.3, and Matplotlib 3.10.1, on a workstation with an Intel Core i9-14900HX processor (Intel, Santa Clara, CA, USA) and 32 GB of RAM; the solver's MIP gap is zero in all reported runs.

Table 2. Model and computational parameters.

Parameter	Symbol	Value
SAA scenarios per day	S	20 (39-bus); 10 (118-bus)
Curtailment penalty	ρ^{w}	500 EUR/MWh
Lost-load penalty	ρ^{s}	9000 EUR/MWh
BESS degradation cost	c^{deg}	5 EUR/MWh
Migration transfer cost	c^{mig}	0.005 EUR/GPU-hour
GPU-hour rental price	c^{gpu}	3 $/GPU-hour (sensitivity 1.25–6.98)
QoS headroom	ϵ	0.02
Migration-arc capacity	$\overline{\phi}$	0.3 of mean workload
DVFS states	(κ_h^P, κ_h^S)	(1.00, 1.00), (0.70, 0.80), (0.45, 0.60)
Checkpoint candidates	$\mathcal{F}$	{2, 4, 7, 9, 17, 35, 70, 140, 280} iterations
CVaR level	α	0.9
Risk weight	λ_{risk}	0.5
Solver	—	Gurobi 13.0.2 (MIP gap 0)

4. Results

4.1. Co-Optimization Against Eight Baselines

Table 3 and Figure 2 summarize the E1 matrix over twelve representative days of 2025. Against the carbon-blind reference B1, the proposed co-optimization reduces system CO_2 emissions by 29.5% (24,185 vs. 17,055 t/day on the 39-bus system at $N = 1.6 \times 10^4$ GPUs per DC) while raising the generation cost by less than 0.01%. Decomposition against the carbon-aware but inflexible reference B2 (17,094 t/day, −29.3%) attributes almost all of the reduction to carbon-priced generation redispatch; the aggregate contribution of all AI-load flexibility channels is a further 0.23% of emissions at this penetration level. Read against the fleet's own footprint rather than system-wide emissions, the instrument carries considerably more weight: the fleet draws ≈627 MWh/day (Section 3), a consumption-weighted ≈152 t/day of CO_2 at the zonal mix of Section 2.5, so the 39 t/day abated by flexibility covers about one quarter (≈26%) of the data centers' own emissions, the unit relevant to an operator deciding whether to expose its load to the dispatch layer. The decomposition also separates the two instruments under evaluation: the carbon signal alone (B2) captures the redispatch wedge, while the scheduling framework contributes the flexibility increment beyond that signal, the grid-compatibility of the endogenous checkpoint policy, and the calibrated risk certificate of Section 4.5. Within that contribution the channels are asymmetric: storage alone (B3, −0.02% vs. B2) captures an order of magnitude less than migration+DVFS alone (B4, −0.22%). At the 118-bus/1k configuration the storage-only case sits 0.3 t/day above the inflexible reference (19,807.7 vs. 19,807.4 t/day in Table 3), a 1.5×10^{-5} relative difference that arises because the BESS is dispatched against the cost signal and its round-trip losses are recharged by generation whose carbon content

differs marginally across hours. B5, which fixes the checkpoint interval at the Daly–Young optimum, matches the proposed method within 0.02% (17,053 vs. 17,055 t/day): the endogenous policy independently selects the reliability-optimal interval ($f = 17$ at every site), so the reliability policy is grid-compatible rather than a competing objective.

Table 3. E1 main results over the twelve representative days; entries are day-means, the CO_2 dispersion is the s.d. across days, and t_{solve} is the mean solve time.

Case/*N*/Baseline	CO_2 [t] (Mean ± s.d.)	Cost [EUR]	Curt. [MWh]	t_{solve} [s]
118/1k/B1 blind	20,467 ± 8172	3,393,081	23.0	14.7
118/1k/B2 no-flex	19,807 ± 8692	3,413,747	23.1	11.3
118/1k/B3 BESS	19,808 ± 8693	3,413,519	23.0	11.9
118/1k/B4 mig+DVFS	19,806 ± 8692	3,413,277	23.1	13.3
118/1k/B5 D–Y	19,806 ± 8693	3,413,048	23.0	13.6
118/1k/B6 signal	20,467 ± 8171	3,393,352	23.1	9.4
118/1k/Proposed	19,806 ± 8693	3,413,053	23.0	11.7
118/16k/B1 blind	20,593 ± 8158	3,421,389	21.0	12.2
118/16k/B2 no-flex	19,923 ± 8682	3,452,895	22.3	6.7
118/16k/B3 BESS	19,928 ± 8689	3,449,290	21.0	7.3
118/16k/B4 mig+DVFS	19,903 ± 8684	3,445,309	22.0	8.7
118/16k/B5 D–Y	19,907 ± 8691	3,441,697	20.6	11.1
118/16k/B6 signal	20,587 ± 8155	3,425,400	22.0	8.5
118/16k/Proposed	19,908 ± 8691	3,442,051	20.6	10.9
118/100k/B1 blind	21,236 ± 8066	3,588,608	8.7	11.3
118/100k/B2 no-flex	20,530 ± 8597	3,677,573	18.1	7.1
118/100k/B3 BESS	20,539 ± 8637	3,657,244	10.3	7.9
118/100k/B4 mig+DVFS	20,416 ± 8617	3,628,409	15.9	8.9
118/100k/B5 D–Y	20,427 ± 8657	3,607,947	7.1	12.4
118/100k/B6 signal	21,222 ± 8050	3,609,027	16.1	7.8
118/100k/Proposed	20,439 ± 8655	3,613,526	7.4	12.0
118/300k/B1 blind	22,426 ± 7739	4,033,767	2.1	14.5
118/300k/B2 no-flex	21,603 ± 8268	4,256,372	8.3	7.8
118/300k/B3 BESS	21,611 ± 8356	4,195,359	3.3	9.2
118/300k/B4 mig+DVFS	21,373 ± 8359	4,095,899	4.3	9.4
118/300k/B5 D–Y	21,378 ± 8457	4,035,305	1.6	15.6
118/300k/B6 signal	22,409 ± 7697	4,079,948	4.7	8.4
118/300k/Proposed	21,427 ± 8436	4,066,709	1.8	14.7
39/1k/B1 blind	24,157 ± 10,423	8,540,272	1031.8	14.5
39/1k/B2 no-flex	16,885 ± 10,368	8,541,050	1033.8	4.4
39/1k/B3 BESS	16,885 ± 10,369	8,540,927	1032.9	4.4
39/1k/B4 mig+DVFS	16,882 ± 10,368	8,540,449	1033.2	10.7
39/1k/B5 D–Y	16,882 ± 10,368	8,540,327	1032.2	12.3
39/1k/B6 signal	24,214 ± 10,436	8,540,449	1032.8	4.6
39/1k/Proposed	16,882 ± 10,368	8,540,333	1032.3	14.0
39/16k/B1 blind	24,185 ± 10,489	8,581,062	987.2	15.9
39/16k/B2 no-flex	17,094 ± 10,438	8,592,454	1012.4	4.1
39/16k/B3 BESS	17,091 ± 10,441	8,590,377	998.2	4.7
39/16k/B4 mig+DVFS	17,056 ± 10,425	8,582,841	1,002.6	10.6
39/16k/B5 D–Y	17,053 ± 10,427	8,580,769	988.5	16.6
39/16k/B6 signal	24,369 ± 10,490	8,583,683	1002.9	4.3
39/16k/B7 temporal	24,562 ± 10,653	8,590,897	1011.4	3.5
39/16k/B8 combined	24,571 ± 10,608	8,590,572	1011.3	3.4
39/16k/Proposed	17,055 ± 10,427	8,581,226	989.0	13.8
39/100k/B1 blind	24,922 ± 10,790	8,814,149	794.0	22.0
39/100k/B2 no-flex	18,290 ± 10,847	8,879,772	900.3	2.8
39/100k/B3 BESS	18,278 ± 10,855	8,868,003	842.6	5.0
39/100k/B4 mig+DVFS	18,047 ± 10,760	8,820,117	845.9	7.8
39/100k/B5 D–Y	18,032 ± 10,763	8,807,582	791.3	20.5
39/100k/B6 signal	25,126 ± 10,832	8,825,295	849.6	4.3
39/100k/Proposed	18,059 ± 10,771	8,814,547	797.0	21.4
39/300k/B1 blind	26,927 ± 11,381	9,378,963	536.4	22.0
39/300k/B2 no-flex	21,278 ± 11,805	9,561,788	734.0	3.4
39/300k/B3 BESS	21,245 ± 11,902	9,520,955	656.2	7.7
39/300k/B4 mig+DVFS	20,472 ± 11,485	9,385,932	594.2	8.9
39/300k/B5 D–Y	20,420 ± 11,565	9,341,617	521.6	25.3
39/300k/B6 signal	27,184 ± 11,491	9,398,471	613.6	4.0
39/300k/Proposed	20,572 ± 11,617	9,379,002	544.5	27.4

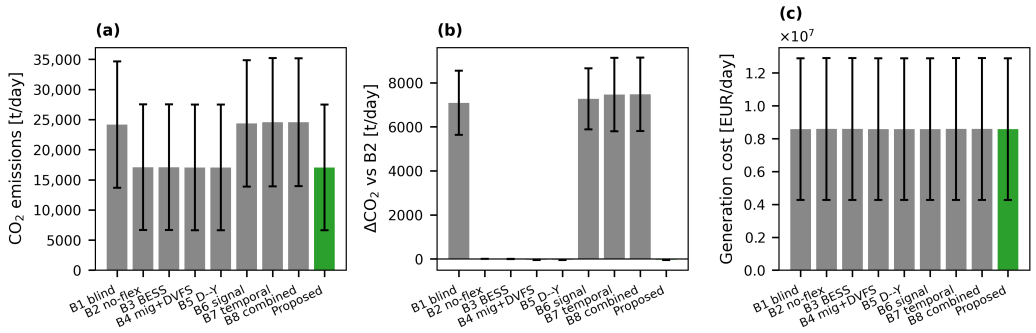


Figure 2. E1 main comparison (39-bus system, $N = 1.6 \times 10^4$, mean ± std over twelve representative days): (**a**) system CO_2 emissions per baseline; (**b**) paired day-level emission difference against the carbon-aware inflexible reference B2 (common random numbers per day); (**c**) generation cost per baseline. In all three panels, green bars denote the proposed method and grey bars denote the baselines.

System-wide renewable curtailment is nearly unchanged (987 vs. 989 MWh/day): migration absorbs spill locally during congested hours, but inter-zonal transmission limits bind in the aggregate (Section 4.2). Load shedding changes little across methods at this scale (scenario-mean $\leq$ 93 MWh/day, $<$0.1% of daily energy, on five days under extreme low-wind scenarios); at $N = 10^5$, where the DC fleet is large enough to matter for adequacy, flexibility cuts worst-day shedding by 29% relative to the inflexible carbon-aware reference (203 vs. 144 MWh/day, Section 4.6).

Baseline B6 isolates the scalar-signal practice: it chases the zonal average carbon-intensity signal (Section 3), moving 98,842 GPU-hours per day into the French zone at $N = 1.6 \times 10^4$ (3.4 times the proposed method's total migration volume, applied at every hour of every day), yet it does not reduce system emissions. On the 39-bus system, B6 sits above the carbon-blind reference B1 by 0.2–1.0% on average across scales (ten of twelve days worse at $N = 1.6 \times 10^4$; up to +2.9% on a single winter day), and on the 118-bus system it is within ±0.1% of B1 at every scale. The mechanism is structural rather than parametric: zonal average intensity is a poor proxy for the system-level marginal effect of a load shift; it ignores marginal emissions, transmission congestion, and concurrent curtailment, so signal following pushes load toward France even in hours when German wind is spilling and the German marginal generator is clean (cf. Section 4.2). The conflict is frequent: 6.5% of the 8,760 h of 2025 settle at negative day-ahead prices, marking hours in which the marginal generator is renewable and the average-intensity ranking misleads, and on the representative day of Section 4.2 the average ranking orders France below Germany in all 24 h while German wind is curtailed in 22 of them. Where the proposed method internalizes these couplings through the dispatch itself, scalar-signal following cannot; uncoordinated carbon chasing by individual data centers can be worse for the system than no response at all. The channel-isolating variants confirm that the deficit is not specific to the spatial channel: at the reference configuration, the temporal-only follower B7 (Section 3) raises system emissions by 1.56% on average and the combined heuristic B8 by 1.59% (cost impact below 0.11% in both cases), so neither time-shifting within a zone nor combining both channels recovers the system-level benefit of dispatch-internal carbon pricing.

4.2. Flexibility Analysis on Representative Day Operation

Figure 3 exposes the mechanism on the representative winter day (25 January 2025, 39-bus system, $N = 1.6 \times 10^4$).

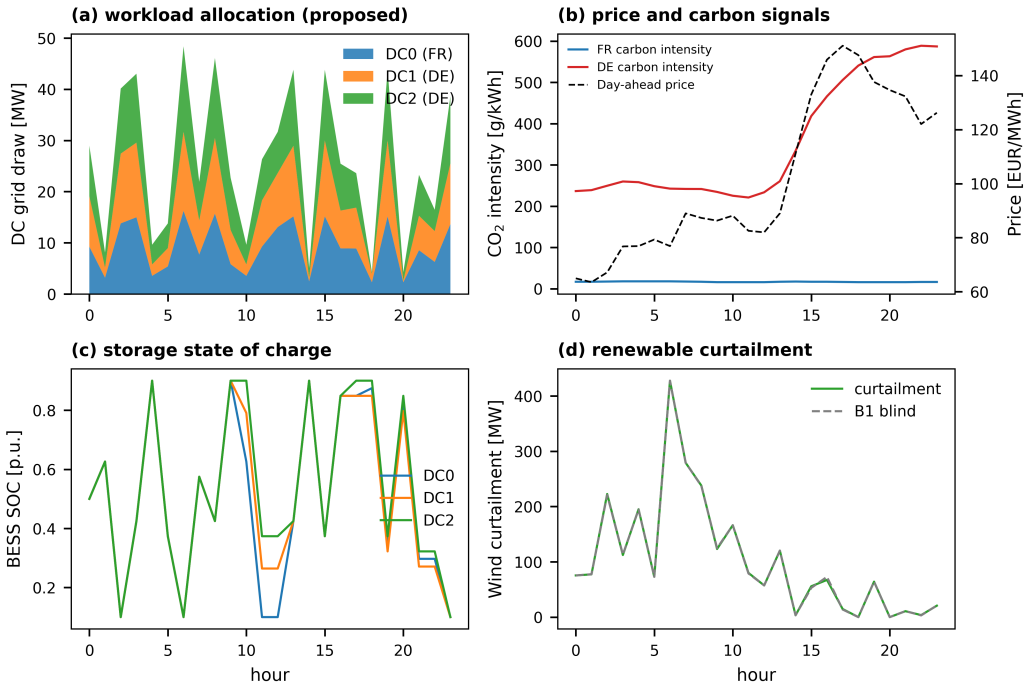


Figure 3. Representative-day operation under the proposed method: (**a**) DC grid draw by site; (**b**) carbon and price signals, namely FR carbon intensity and DE carbon intensity on the left axis with the day-ahead price on the right axis; (**c**) BESS state of charge; (**d**) wind curtailment against the carbon-blind baseline.

The dominant action is on the generation side: the low-carbon French fleet's share of daily energy rises from 30.0% under B1 to 52.7% under the proposed method, a 21.7 GWh/day inter-zonal redispatch that the zone-mean-equalized merit order leaves untapped while carbon is unpriced (the French fleet runs at a 49% daily capacity factor under B1 and 87% once carbon is priced). Workload geography follows a logic of its own: migration is Germany-ward under both methods (≈30k GPU-hours/day) and the French share of DC load is identical (0.354). The flexible component concentrates in hours 0–13, when German wind curtailment reaches 57–428 MW and the German carbon intensity sits at its daily minimum (221–260 g/kWh); during the evening carbon peak (hours 14–23, I_{DE} rising from 334 to 587 g/kWh) curtailment vanishes and migration does not intensify. Flexible load is thus dispatched to absorb congested renewable energy, not to chase the low-carbon zone, and the migration-arc capacity never binds. Migration stays Germany-ward under carbon pricing, and the marginal logic explains why: in curtailment hours the German marginal generator is wind with zero marginal emissions, so absorbing load in Germany is system-cleaner than the zonal average intensities suggest. The BESS fleet performs shallow hour-scale cycling (≈0.3 GWh/day aggregate throughput, symmetric across sites) that smooths the interaction of migration and DVFS actions with unit-commitment constraints, rather than classical price arbitrage. The endogenous checkpoint policy selects $f = 17$ iterations at all three sites, the candidate in $\mathcal{F}$ nearest the exact optimum of Equation (2), at no reliability-cost premium.

4.3. Pareto Analysis on the Price of Carbon

Figure 4 traces the cost–emission front as the incremental carbon price $\pi_{\mathrm{CO_2}}$ (Section 2.5) sweeps 0–200 EUR/t in seven steps $\{0, 20, 50, 70, 100, 150, 200\}$ on the 39-bus system at the reference scale $N = 1.6 \times 10^4$, evaluated over four representative days (25 January, 18 February, 25 March, and 12 April 2025).

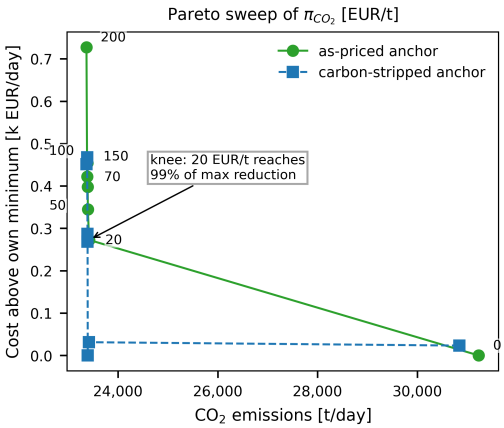


Figure 4. E2 Pareto front of the incremental carbon price π_{CO_2} (point labels in EUR/t) under the as-priced and carbon-stripped cost anchors (Section 2.5); each curve's cost is measured above its own minimum.

The front is steep at the origin and flattens thereafter: an incremental price of 20 EUR/t already achieves a 25.0% system emission reduction (from 31,230 to 23,421 t/day on average), i.e., 99% of the maximum reduction attainable at 200 EUR/t (25.2%), at a generation-cost increase below 0.05%. The redispatch toward the low-carbon fleet raises generation cost by less than 0.05% under the equalized zone-mean merit order. Beyond about 50 EUR/t the front is nearly flat: the remaining emissions stem from DE-zone generation that cannot be displaced within the FR fleet's capacity and the network's transmission limits. The stripped-anchor variant of Section 2.5 removes the implicit carbon signal embedded in observed market prices (a -26% shift of the energy-cost level), thereby testing whether the knee of the front is an artifact of double-counted emission costs. Its front deviates from the unmodified anchor by at most 1.3% in emissions at $\pi_{CO_2} = 0$ and by less than 0.01% for $\pi_{CO_2} \geq 20$ EUR/t, with the knee unchanged: the front characterizes the marginal effect of strengthening the carbon signal above its market-embedded level, which informs the policy discussion in Section 5.

4.4. Ablation Studies

Table 4 and Figure 5 report the E3 ablation on a representative winter day, evaluated with common random numbers so that differences are attributable to the treatment alone.

Table 4. E3 ablation results.

Setting	CO_2 [t]	Cost [EUR]	Curt. [MWh]
band = conformal	17,282	9,949,939	2028.0
band = empirical	17,290	9,953,475	2021.7
bess_scale = 0.5	17,283	9,950,658	2040.7
bess_scale = 1.0	17,282	9,949,939	2028.0
bess_scale = 1.5	17,279	9,949,312	2015.9
lambda_risk = 0.0	17,279	9,949,886	2025.8
lambda_risk = 0.5	17,282	9,949,939	2028.0
lambda_risk = 1.0	17,282	9,949,968	2027.9
lambda_risk = 2.0	17,282	9,949,999	2027.9
mig_cap_frac = 0.15	17,282	9,949,939	2028.0
mig_cap_frac = 0.3	17,282	9,949,939	2028.0
mig_cap_frac = 0.6	17,282	9,949,939	2028.0

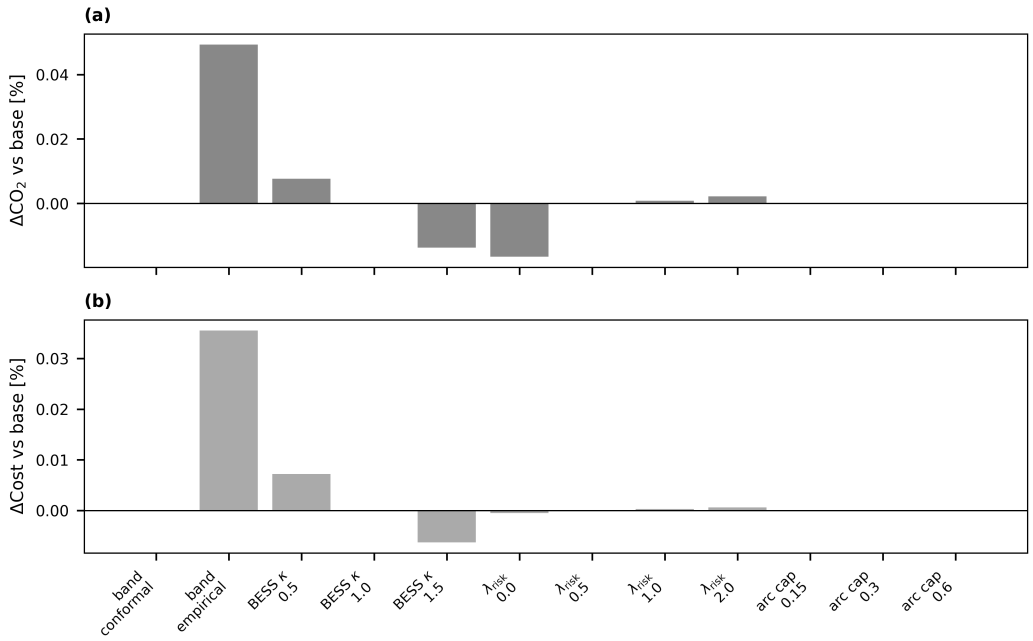


Figure 5. E3 ablation of calibration method, risk aversion, migration capacity, and BESS sizing, shown as deviation from the shared base setting (conformal band, $\lambda_{\text{risk}} = 0.5$, base arc capacity, base BESS sizing): (**a**) emissions; (**b**) cost.

Across these sweeps, little moves: sweeping risk aversion λ_{risk} over $\{0, 0.5, 1, 2\}$ and migration-arc capacity over $\{0.15, 0.3, 0.6\}$ changes expected emissions and cost by less than 0.02%, since the conformal-clipped scenario bands keep scenario costs close enough that tail-risk weighting barely moves the first stage, and the base migration capacity is not binding on this day. BESS sizing over $\kappa \in \{0.5, 1.5\}$ spans a 0.02% emission range. Switching from empirical to conformal bands lowers expected cost by 0.04% (9.950 vs. 9.954 M EUR/day) and emissions by 0.05%; the benefit of the calibrated bands lies in the tail rather than the mean, as the coverage results of Section 4.5 show.

4.5. Out-of-Sample Validation of the Certificate

Table 5 and Figure 6 report the out-of-sample validation (E4).

Table 5. E4 out-of-sample certificate coverage: same-day fresh scenarios (200 draws per training day, the conformal guarantee domain) versus cross-day transfer to 500 held-out realized days of 2025 (the n column counts the cross-day evaluation).

Band	Same-Day Cov.	Cross-Day Cov.	Nominal	Infeasible	n
Conformal	0.975	0.802	0.9	0	500
Empirical	0.860	0.778	0.9	0	500

With the first-stage decisions fixed from the representative winter day, the in-sample $\alpha = 0.9$ cost certificate is evaluated on two held-out domains. One evaluation sample is a single recourse solve under one exogenous series with the first stage frozen: for the same-day domain, 200 fresh unclipped residual perturbations of the training day; for the cross-day domain, 500 draws with replacement from the 364 remaining realized days of 2025, each contributing its realized price, wind, load, and carbon series. A sample counts as covered when its realized cost lies within a 5% relative tolerance above the certificate, absorbing solver and sampling noise. On the same-day domain (the domain of the formal conformal marginal guarantee), the conformal-band certificate covers 97.5% of realized costs (Wilson 95% interval [94.3, 98.9]%), while the empirical-band certificate under-covers

at 86.0% [80.5, 90.1]% against the 90% nominal level. The conformal certificate is 8.7% higher (13.8 vs. 12.7 M EUR). On the harder cross-day transfer test (an empirical diagnostic outside the guarantee's domain), coverage falls to 80.2% [76.5, 83.5]% (conformal) versus 77.8% [74.0, 81.2]% (empirical), a direct measure of the seasonal distribution shift, and recourse stays feasible on every evaluated day. Realized cross-day emissions average 17.7 kt/day, within 3% of the in-sample estimate. Operationally, first-stage decisions applied to future days therefore call for a rolling certificate: re-solving the first stage daily and updating the conformal quantile with recent residuals, in the spirit of adaptive conformal calibration [56].

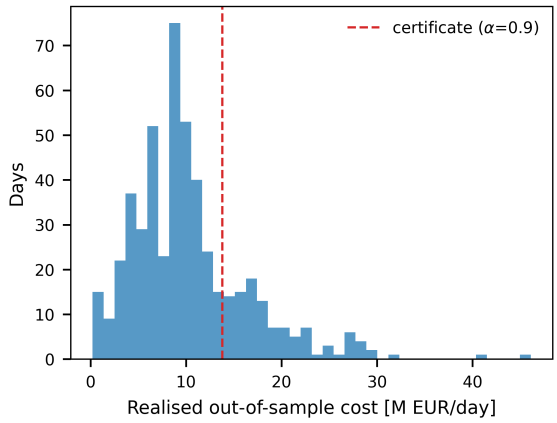


Figure 6. E4 out-of-sample realized-cost distribution with the conformal certificate at $\alpha = 0.9$ (dashed).

4.6. Scale Effects

Figure 7 reports the scale analysis across $N \in \{10^3, 1.6 \times 10^4, 10^5, 3 \times 10^5\}$ GPUs per DC on both test systems, spanning DC-fleet shares of system load from 0.04% to 11.3%.

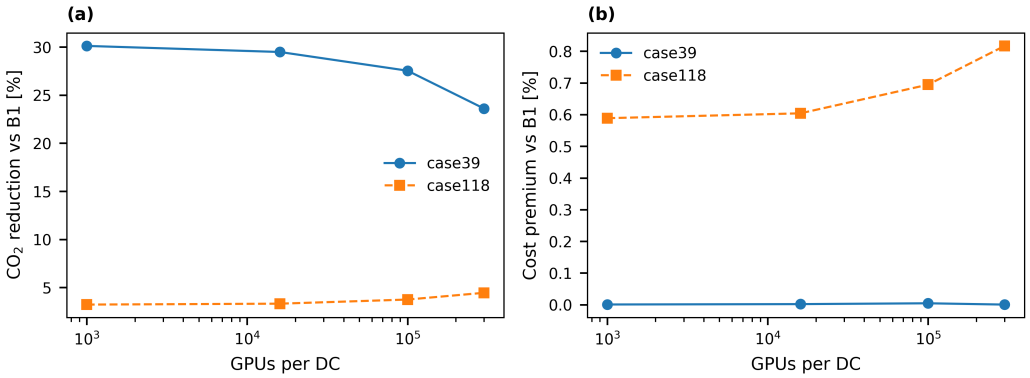


Figure 7. E5 scale effects: (**a**) system CO_2 reduction and (**b**) generation-cost premium of the proposed method relative to the carbon-blind baseline, versus GPU fleet size per DC on both test systems.

All scale conclusions rest on the full twelve-day E1 matrix; the proposed configuration is additionally replicated day by day in E5 (twelve representative days, both systems, Section 3), reproducing the E1 values exactly. Two opposing trends emerge as penetration grows. The total reduction relative to the carbon-blind baseline narrows from 30.1% to 23.6% on the 39-bus system, because the added DC load must itself be served by a fleet whose low-carbon headroom is finite. From a 0.02% foothold at $N = 10^3$, the flexibility channels grow super-linearly: their aggregate value beyond the inflexible reference rises through 0.23% ($N = 1.6 \times 10^4$) and 1.3% ($N = 10^5$) to 3.3% ($N = 3 \times 10^5$) of emissions, with migration volume reaching 5.9×10^5 GPU-hours/day. Relative to the fleet's own consumption-weighted footprint, the same channels abate a roughly scale-invariant one quarter (24–26% from $N = 1.6 \times 10^4$ to 3×10^5): flexibility removes a stable slice of the

AI load's own emissions while the system-level share catches up as penetration grows. Flexibility also becomes an adequacy resource at scale: worst-day scenario-mean load shedding falls from 849 to 431 MWh (−49%) at $N = 3 \times 10^5$ relative to the inflexible carbon-aware reference, a benefit that is absent at $N \leq 1.6 \times 10^4$ where shedding is method-independent. All shedding figures reported in this paper are scenario means, and worst-day values are the maxima of these means across the twelve days. The benefit extends across the scenario distribution, not only in the mean: on that worst day (18 February 2025) the per-scenario 95th-percentile shed falls from 1877 to 878 MWh and the worst single scenario from 1904 to 921 MWh. The 118-bus system shows the same trends at lower levels (total savings 3.2–4.5%; flexibility value up to 0.8%), confirming that topology changes the magnitude, not the direction, of the effect.

One exception occurs: at $N = 3 \times 10^5$ the proposed method sits above the fixed Daly–Young reference on both test systems (by 0.74% on the 39-bus system, 20,572 vs. 20,420 t/day, and by 0.23% on the 118-bus system, 21,427 vs. 21,378 t/day), the only configurations in which the ranking inverts. Two features of the model price the gap: the checkpoint-overhead term in the QoS margin (8) is carried only when f is endogenous, and at this scale the migration choice is near-degenerate, its transfer cost ($\approx$2.5 kEUR/day against a 4 M EUR/day objective on the 118-bus system) too small to break ties between near-equal schedules; within its own constraint set the joint model dominates the fixed-f variant by construction. Applying the overhead term of Equation (8) uniformly to all methods, including baselines with an exogenously fixed interval, removes the asymmetry (a one-line change in the released code), and since the retained conservative form biases the comparison against the proposed method, the reported flexibility values are lower bounds; evaluating the harmonized variant is left to future work. Solve times remain tractable throughout (10.0 s median, 51.4 s worst case across the 672 paired runs of the main E1 matrix, with the solver's MIP gap at zero).

4.7. Sensitivity to Structural Model Choices

Two structural choices underpin the case study: the FR/DE fleet assignment and the training-failure model. We test both.

4.7.1. Fleet-Assignment Sensitivity

The baseline mapping assigns the cheapest 40% of generation capacity to the French fleet (Section 2.1). Re-running the E1 comparison (B1, B2, and the proposed method; 39-bus system, $N = 1.6 \times 10^4$, twelve representative days, common random numbers) with shares of 30% and 50% yields total reductions of 18.9% and 43.1% relative to B1 (vs. 29.5% at the baseline share), with the flexibility increment beyond B2 at 0.17%–0.32% (vs. 0.23%). Absolute emission levels move with the assignment (B1 spans 27.6–20.2 kt/day across the three shares), and the carbon-signal headroom grows with the low-carbon fleet's size, while the direction of every reported effect and the ranking of all methods are unchanged.

4.7.2. Failure-Model Sensitivity

The checkpoint economics depend on the per-GPU failure rate λ and the cascade factor γ_{cascade} (Section 2.2). Scaling λ over $\{0.25, 0.5, 1, 2, 4\}\times$ its calibrated value and testing $\gamma_{\text{cascade}} \in \{0, 0.32\}$ on two representative days (25 January and 2 July 2025, 39-bus system, $N = 1.6 \times 10^4$) leaves every scheduling conclusion in place: the endogenous checkpoint choice tracks the Daly–Young scaling throughout (f shifts from 35 to 9 iterations as λ spans its sixteen-fold range and stays at 17 for $\gamma_{\text{cascade}} \in \{0, 0.32\}$, in every configuration the candidate nearest the exact optimum of Equation (2)), and system emissions and cost vary by less than 0.02% relative to the calibrated model. The reliability-optimal interval thus stays grid-compatible well beyond the calibrated failure assumptions.

5. Discussion

5.1. Principal Findings

Several results carry beyond the case study. Pricing carbon in dispatch reduces system CO_2 emissions by 29.5% at the reference 1.6×10^4-GPU scale (23.6–30.1% across the tested scales) relative to carbon-blind dispatch on the 39-bus system (3.3–4.5% on the 118-bus system, whose carbon-blind dispatch already engages most of the low-carbon fleet's headroom) at a generation-cost increase below 0.01% (0.6–0.8% on the 118-bus system). The decomposition is unambiguous: carbon-priced generation redispatch delivers almost all of this reduction, an instrument available to the operator irrespective of data-center cooperation. For a grid hosting AI loads, explicit carbon pricing of dispatch thus offers a larger lever than any hardware-side measure. The AI-load channels look different depending on the denominator: a further 0.23% of system emissions at current penetration, about one quarter of the fleet's own consumption-weighted footprint at every tested scale, and an adequacy resource at gigawatt scale (Sections 4.1 and 4.6). The two denominators answer different questions (system share for the planner, fleet footprint for the operator), and keeping them apart is the case for exposing AI loads to the dispatch layer.

Prevailing practice in carbon-aware computing fares worse than coordination. Chasing the zonal average carbon intensity, as hyperscalers do today, raises system emissions relative even to carbon-blind co-optimized dispatch (Section 4.1), because average-signal migration ignores transmission limits and marginal redispatch effects.

Reliability and grid objectives turn out to be compatible rather than competing: the endogenous checkpoint policy independently settles on the Daly–Young optimum in every tested configuration. From today's 0.02% foothold, the flexibility channels grow super-linearly toward the gigawatt-campus era projected for 2030 (Section 4.6); their eventual value will be shaped by the siting and market-design choices made in the near term.

On the uncertainty side, the split-conformal scenario bands meet their nominal coverage out of sample on the guarantee's own domain, while uncalibrated empirical bands under-cover (Section 4.5); under cross-day distribution shift both degrade in a quantified way, with the conformal band ahead throughout, and recourse remains feasible on all 500 evaluated realized days.

5.2. Interpretation

The dominant mechanism behind the emission savings is generation redispatch, not load movement: pricing carbon at the dispatch layer re-ranks the FR and DE fleets within the network's transmission limits, and the wide carbon-intensity contrast between the nuclear-dominated French and the coal-exposed German system (Section 2.5) makes the redispatch highly effective.

This magnitude is specific to the constructed equal-merit mapping: in the coupled market as observed, the price merit order already favors low-cost French nuclear, so part of this redispatch is delivered by prices alone and the additional saving available from carbon pricing is smaller. The 29.5% figure is accordingly the carbon-signal headroom of the mapped test system, not a forecast of achievable savings in the observed market: all headline reductions are statements about the mapped benchmark under the fleet-assignment convention of Section 2.1, which Section 4.7 varies. Workload migration, by contrast, is used opportunistically, toward congested regions where wind would otherwise be curtailed and flexible DC load can absorb the spill, rather than as a systematic move to the low-carbon zone. Two implications follow for the carbon-aware-computing literature. Temporal shifting within a single grid, the mechanism most operators deploy today, reaches a fraction of the available reduction, since the large wedge is inter-zonal and sits on the generation side. Scalar-signal following, beyond that, exposes a network externality that

dispatch-internal pricing avoids: migration toward a low-carbon zone relieves congestion and curtailment up to transmission limits, beyond which marginal emissions can move against the average-intensity signal; the co-optimized schedule respects these limits by construction.

A natural market-design question is whether exposing data centers to marginal rather than average carbon signals (locational marginal carbon intensity or dynamic nodal carbon tariffs) would close the gap between autonomous signal following and co-optimized dispatch. Within our model the answer is affirmative: the emission-priced dispatch internalizes exactly the marginal redispatch effects that the average signal misses, so a nodal marginal-carbon tariff would approximate the co-optimized outcome without requiring data-center cooperation. Practical barriers stand between that equivalence and deployment: marginal signals are considerably more volatile than zonal averages, their computation requires settlement-grade marginal-emissions accounting that today's markets do not publish, and a price signal alone still cannot coordinate the workload's reliability policy and network-constrained migration the way the joint formulation does. Designing and testing such tariffs is a promising direction for future work.

Signal following fares the same across the three channels of prevailing practice: spatial migration (B6), temporal shifting (B7), and their combination (B8) all raise emissions relative to carbon-blind co-optimized dispatch (Section 4.1), so the externality is structural rather than channel-specific. Within the co-optimized model, the temporal channel is exercised through DVFS (Section 4.6).

5.3. Sustainability Implications

For data-center operators, the results quantify the system value of exposing reliability policy, migration, and storage to the dispatch layer: the Pareto sweep shows that 99% of the achievable emission reduction is already attained at an incremental carbon price of 20 EUR/t on top of the allowance costs that market prices already internalize (Section 2.5), so even a modest strengthening of the carbon signal beyond its market-embedded level captures nearly all of the benefit, and the stripped-anchor sensitivity of Section 4.3 confirms the knee. The signal-following result is equally actionable: autonomous response by individual data centers can leave the system worse off than no response (Section 4.1), which argues for integrating large AI loads into dispatch coordination (or exposing them to carbon-priced markets) rather than leaving them to autonomous signal response. For system operators and regulators, the finding that a small number of large, cooperative AI loads can materially reduce curtailment and improve adequacy at the largest tested fleets (Section 4.6) argues for treating hyperscale campuses as flexibility assets in resource adequacy and network planning rather than as passive demand. For the research community, the fully reproducible pipeline lowers the barrier to replicating and extending system-level studies of AI-load integration.

5.4. Limitations

Five modeling choices bound the scope of the claims. (i) The zonal mapping onto benchmark networks is a reduced form: it preserves network physics and empirical signals but not the true European topology, so absolute magnitudes, not directions or relative rankings, carry modeling error. (ii) The DE carbon intensity is computed from generation-weighted average emission factors and omits minor fuel categories; this slightly overstates intensity, so the reported savings err on the conservative side. Average rather than marginal emissions are used throughout, and marginal signals could shift the ranking of hours if not of zones. (iii) Day-ahead prices are assumed known at scheduling time and common across zones, so spatial price arbitrage, which in reality would partly align with the carbon signal

(French nuclear is both cheap and clean), is absent, and the migration-cost sensitivity covers it only partially. (iv) The thermal model uses static PUE and the storage model a constant efficiency with linear throughput degradation. A weather-dependent PUE would couple the cooling overhead to ambient temperature and thereby to the same solar cycle that shapes the midday carbon dip: hotter afternoons raise the DC draw precisely when solar suppresses carbon intensity, partially offsetting the value of temporal flexibility, whereas cold, dark winter peaks would amplify it; this co-variation could re-rank the hours in which flexibility is most valuable, though not the inter-zonal redispatch wedge that dominates the results. Cycle-depth- and temperature-dependent battery aging would likewise raise the effective cost of deep throughput relative to the shallow cycling observed here. (v) Marginal costs are anchored to observed market prices that internalize EU ETS allowance costs; π_{CO_2} is therefore an incremental shadow price above the market-embedded level, and absolute cost levels inherit this embedded-carbon convention, while the redispatch ranking does not (Section 4.3).

6. Conclusions

We have treated the sustainable grid integration of AI training data centers as what it physically is: a co-optimization problem in which the workload's own reliability policy becomes a dispatchable, carbon-priced decision. The resulting two-stage, risk-averse model couples unit commitment and network physics with endogenous checkpoint-interval selection, geographical workload migration, DVFS, and behind-the-meter storage, driven by split-conformal-calibrated scenarios built from a full year of public French and German system data. On IEEE 39- and 118-bus test systems, carbon-priced dispatch cuts system CO_2 emissions by 29.5% at the reference scale (up to 30.1% at 10^3 GPUs per site) relative to carbon-blind dispatch, at a generation-cost premium below 0.01% (0.6–0.8% on the 118-bus system), and the decomposition assigns more than 99% of it at reference penetration to generation redispatch. The AI-load flexibility channels contribute a further 0.23% today (about one quarter of the fleet's own carbon footprint, a proportion roughly constant across scales) and grow super-linearly to 3.3% at gigawatt scale, where they also improve adequacy (up to 49% less worst-day load shedding). The endogenous checkpoint policy independently recovers the Daly–Young optimum in every tested configuration: reliability-optimal training is already grid-compatible. Following the scalar average-intensity signal—the prevailing practice in carbon-aware computing—can instead raise system-level emissions even relative to carbon-blind co-optimized dispatch. Split-conformal calibration maintains out-of-sample certificate coverage at or above its nominal level, a guarantee uncalibrated scenario bands do not provide. Cooperative AI loads thus emerge as a measurable asset for SDG 7 and SDG 13 objectives, and the reproducible, open-data pipeline gives operators, planners, and regulators a concrete basis for integrating the AI build-out into a decarbonizing power system.

Author Contributions: Conceptualization, J.Z. and Y.W.; methodology, J.Z.; software, J.Z.; validation, J.Z.; formal analysis, J.Z.; investigation, J.Z.; data curation, J.Z.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z. and Y.W.; visualization, J.Z.; supervision, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All source data used in this study are derived from publicly available, token-free sources (RTE éCO2mix [61], SMARD [62], MATPOWER [57]); details and access links are provided in Section 2. The processed datasets and the code used to generate the results are available from the corresponding author on reasonable request.

Acknowledgments: The authors would like to acknowledge the use of Gemini 3.5 for language polishing, editing, and grammar enhancement of the manuscript. The AI tool was not used for any research-related operations. All original content of the manuscript was independently completed by the authors, who bear full responsibility for the originality and scientificity of all research content and conclusions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
BESS	Battery energy storage system
CVaR	Conditional value at risk
DC	Data center
DVFS	Dynamic voltage and frequency scaling
EU ETS	European Union Emissions Trading System
MILP	Mixed-integer linear program
PUE	Power usage effectiveness
QoS	Quality of service
SAA	Sample average approximation
SDG	Sustainable Development Goal
UC	Unit commitment

References

1. International Energy Agency. *Electricity 2024: Analysis and Forecast to 2026*; Technical Report; International Energy Agency: Paris, France, 2024; IEA World Energy Outlook series report.
2. International Energy Agency. *Energy and AI*; Technical Report; International Energy Agency: Paris, France, 2025; World Energy Outlook Special Report.
3. Shehabi, A.; Smith, S.; Sartor, D.; Brown, R.; Herrlin, M.; Koomey, J.; Masanet, E.; Horner, N.; Azevedo, I.; Lintner, W. *United States Data Center Energy Usage Report*; Technical Report LBNL-1005775; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2016.
4. Masanet, E.; Shehabi, A.; Lei, N.; Smith, S.; Koomey, J. Recalibrating Global Data-Center Energy-Use Estimates. *Science* **2020**, *367*, 984–986. https://doi.org/10.1126/science.aba3758.
5. Patterson, D.; Gonzalez, J.; Le, Q.; Liang, C.; Munguia, L.M.; Rothchild, D.; So, D.; Texier, M.; Dean, J. Carbon Emissions and Large Neural Network Training. *arXiv* **2021**, arXiv:2104.10350 .
6. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *Commun. ACM* **2020**, *63*, 54–63. https://doi.org/10.1145/3381831.
7. Wu, C.J.; Raghavendra, R.; Gupta, U.; Acun, B.; Ardalani, N.; Maeng, K.; Chang, G.; Aga, F.; Huang, J.; Bai, C.; et al. Sustainable AI: Environmental Implications, Challenges and Opportunities. *Proc. Mach. Learn. Syst.* **2022**, *4*, 795–813.
8. Kaack, L.H.; Donti, P.L.; Strubell, E.; Kamiya, G.; Creutzig, F.; Rolnick, D. Aligning Artificial Intelligence with Climate Change Mitigation. *Nat. Clim. Change* **2022**, *12*, 518–527. https://doi.org/10.1038/s41558-022-01377-7.
9. Google LLC. *Moving Toward 24×7 Carbon-Free Energy at Google Data Centers: Progress and Insights*; Technical Report; Google LLC: Mountain View, CA, USA, 2018.
10. Manganelli, M.; Soldati, A.; Martirano, L.; Ramakrishna, S. Strategies for Improving the Sustainability of Data Centers via Energy Mix, Energy Conservation, and Circular Energy. *Sustainability* **2021**, *13*, 6114. https://doi.org/10.3390/su13116114.
11. Mondal, S.; Bin Faruk, F.; Rajbongshi, D.; Efaz, M.M.K.; Islam, M.M. GEECO: Green Data Centers for Energy Optimization and Carbon Footprint Reduction. *Sustainability* **2023**, *15*, 15249. https://doi.org/10.3390/su152115249.
12. Chountalas, P.T.; Chatzoglou, P.; Vlachos, D. Modeling Critical Success Factors for Green Energy Integration in Data Centers. *Sustainability* **2025**, *17*, 3543. https://doi.org/10.3390/su17083543.

13. Durmus Senyapar, H.N.; Bayindir, R. The Energy Hunger Paradox of Artificial Intelligence: End of Clean Energy or Magic Wand for Sustainability? *Sustainability* **2025**, *17*, 2887. https://doi.org/10.3390/su17072887.
14. Penneigh, R.D. Biomass Power Generation and Energy Management in Smart Grid-Connected Data Centers: A Comprehensive Review and Alignment Framework. *Sustainability* **2026**, *18*, 6141. https://doi.org/10.3390/su18126141.
15. Cioara, T.; Antal, M.; Antal, C.D.; Anghel, I.; Bertoncini, M.; Arnone, D.; Lazzaro, M.; Mammina, M.; Velivassaki, T.H.; Voulkidis, A.; et al. Data Centers Optimized Integration with Multi-Energy Grids: Test Cases and Results in Operational Environment. *Sustainability* **2020**, *12*, 9893. https://doi.org/10.3390/su12239893.
16. Li, Q.; Wang, F.; Wu, J.; Li, Q.; Zhang, G. Multivariate Coupling Model and Reservoir Characteristics of Enhanced Geothermal Reservoirs. *Energies* **2026**, *19*, 3180. https://doi.org/10.3390/en19133180.
17. Ao, F.; Li, Q.; Li, Q.; Wang, J.; Wang, F.; Yan, C. Numerical Simulation Investigation of Fracture Propagation Behavior Patterns and Sensitivity Factors of Oil Shale Reservoirs in the Xunyi Region Considering the Influence of Natural Fracture. *Geofluids* **2025**, *2025*, 2762142. https://doi.org/10.1155/gfl/2762142.
18. Wu, J.; Li, Q.; Li, Q.; Wang, F.; Cheng, Y.; Yan, C. Design and Optimization of Bottom-Hole Temperature–Pressure Combinations in Gas Production from Gas Hydrates via Carbon Dioxide Replacement Strategy. *Energies* **2026**, *19*, 3536. https://doi.org/10.3390/en19153536.
19. Qureshi, A.; Weber, R.; Balakrishnan, H.; Guttag, J.; Maggs, B. Cutting the Electric Bill for Internet-Scale Systems. *ACM SIGCOMM Comput. Commun. Rev.* **2009**, *39*, 123–134. https://doi.org/10.1145/1594977.1592584.
20. Liu, Z.; Lin, M.; Wierman, A.; Low, S.H.; Andrew, L.L.H. Greening Geographical Load Balancing. *ACM SIGMETRICS Perform. Eval. Rev.* **2011**, *39*, 193–204. https://doi.org/10.1145/2007116.2007139.
21. Lin, M.; Liu, Z.; Wierman, A.; Andrew, L.L.H. Online Algorithms for Geographical Load Balancing. In *Proceedings of the 2012 International Green Computing Conference (IGCC), San Jose, CA, USA, 4–8 June 2012*; IEEE: New York, NY, USA, 2012; pp. 1–10. https://doi.org/10.1109/IGCC.2012.6322266.
22. Lin, M.; Wierman, A.; Andrew, L.L.H.; Thereska, E. Dynamic Right-Sizing for Power-Proportional Data Centers. *IEEE/ACM Trans. Netw.* **2013**, *21*, 1378–1391. https://doi.org/10.1109/TNET.2012.2226216.
23. Liu, Z.; Wierman, A.; Chen, Y.; Razon, B.; Chen, N. Data Center Demand Response: Avoiding the Coincident Peak via Workload Shifting and Local Generation. *Perform. Eval.* **2013**, *70*, 770–791. https://doi.org/10.1016/j.peva.2013.08.014.
24. Ghatikar, G. *Demand Response Opportunities and Enabling Technologies for Data Centers: Findings from Field Studies*; Technical Report LBNL-5763E; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2012.
25. Wang, H.; Huang, J.; Lin, X.; Mohsenian-Rad, H. Proactive Demand Response for Data Centers: A Win-Win Solution. *IEEE Trans. Smart Grid* **2015**, *7*, 1584–1596. https://doi.org/10.1109/tsg.2015.2501808.
26. Ghamkhari, M.; Mohsenian-Rad, H. Data Centers to Offer Ancillary Services. In *Proceedings of the 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm), Tainan, Taiwan, 5–8 November 2012*; IEEE: New York, NY, USA, 2012; pp. 436–441. https://doi.org/10.1109/SmartGridComm.2012.6486023.
27. Aikema, D.; Simmonds, R.; Zareipour, H. Data Centres in the Ancillary Services Market. In *Proceedings of the 2012 International Green Computing Conference (IGCC), San Jose, CA, USA, 4–8 June 2012*; IEEE: New York, NY, USA, 2012; pp. 1–10. https://doi.org/10.1109/IGCC.2012.6322252.
28. Aksanli, B.; Rosing, T.S. Providing Regulation Services and Managing Data Center Peak Power Budgets. In *Proceedings of the 2014 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, Germany, 24–28 March 2014*; IEEE: New York, NY, USA, 2014, pp. 1–6. https://doi.org/10.7873/DATE.2014.156.
29. Radovanovic, A.; Koningstein, R.; Schneider, I.; Chen, B.; Duarte, A.; Roy, B.; Xiao, D.; Haridasan, M.; Hung, P.; Care, N.; et al. Carbon-Aware Computing for Datacenters. *IEEE Trans. Power Syst.* **2023**, *38*, 1270–1280. https://doi.org/10.1109/TPWRS.2022.3173250.
30. Google LLC. Our Data Centers Now Work Harder When the Sun Shines and Wind Blows. In *The Keyword (Google Blog)*; Google LLC: Mountain View, CA, USA, 2020.
31. Li, B.; Samsi, S.; Gadepally, V.; Tiwari, D. Clover: Toward Sustainable AI with Carbon-Aware Machine Learning Inference Service. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'23), Denver, CO, USA, 12–17 November 2023* ; ACM: New York, NY, USA, 2023; pp. 1–15. https://doi.org/10.1145/3581784.3607034.
32. Hjalmarsson, J.; Thomas, K.; Boström, C. Service Stacking Using Energy Storage Systems for Grid Applications—A Review. *J. Energy Storage* **2023**, *60*, 106639. https://doi.org/10.1016/j.est.2023.106639.
33. Elshurafa, A.M.; AlCulail, N.; AlMuhanna, A.; Amer, M. The Value of Storage in Electricity Generation: A Qualitative and Quantitative Review. *J. Energy Storage* **2020**, *32*, 101872. https://doi.org/10.1016/j.est.2020.101872.
34. Ratshitanga, M.; Ayeleso, A.; Krishnamurthy, S.; Rose, G.; Aminou Moussavou, A.A.; Adonis, M. Battery Storage Use in the Value Chain of Power Systems. *Energies* **2024**, *17*, 921. https://doi.org/10.3390/en17040921.

35. Bian, N.; Kawasaki, S. Two Stage Stochastic Optimization Scheduling of Power System Considering Thermal Power Units and Energy Storage Peak Shaving Pricing Strategies. *J. Electr. Eng. Technol.* **2025**, *20*, 2961–2975. https://doi.org/10.1007/s42835-025-02200-x.
36. Arbabzadeh, M.; Johnson, J.X.; Keoleian, G.A.; Rasmussen, P.G.; Thompson, L.T. The Role of Energy Storage in Deep Decarbonization of Electricity Production. *Nat. Commun.* **2019**, *10*, 3413. https://doi.org/10.1038/s41467-019-11161-5. Correction in *Nat. Commun.* **2019**, *10*, 3909.
37. Young, J.W. A First Order Approximation to the Optimum Checkpoint Interval. *Commun. ACM* **1974**, *17*, 530–531. https://doi.org/10.1145/361147.361115.
38. Daly, J.T. A Higher Order Estimate of the Optimum Checkpoint Interval for Restart Dumps. *Future Gener. Comput. Syst.* **2006**, *22*, 303–312. https://doi.org/10.1016/j.future.2004.11.016.
39. Cappello, F.; Geist, A.; Gropp, W.; Kale, L.; Kramer, B.; Snir, M. Toward Exascale Resilience: 2014 Update. *Supercomput. Front. Innov.* **2014**, *1*, 5–28. https://doi.org/10.14529/jsfi140101.
40. Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. The Llama 3 Herd of Models. *arXiv* **2024**, arXiv:2407.21783 .
41. Jiang, Z.; Lin, H.; Zhong, Y.; Huang, Q.; Chen, Y.; Zhang, Z.; Peng, Y.; Li, X.; Xie, C.; Nong, S.; et al. MegaScale: Scaling Large Language Model Training to More Than 10,000 GPUs. In *Proceedings of the 21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24), Santa Clara, CA, USA, 16–18 April 2024* ; USENIX Association: Berkeley, CA, USA, 2024; pp. 745–764.
42. Jiang, R.; Wang, J.; Guan, Y. Robust Unit Commitment with Wind Power and Pumped Storage Hydro. *IEEE Trans. Power Syst.* **2012**, *27*, 800–810. https://doi.org/10.1109/TPWRS.2011.2169817.
43. Bertsimas, D.; Litvinov, E.; Sun, X.A.; Zhao, J.; Zheng, T. Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem. *IEEE Trans. Power Syst.* **2013**, *28*, 52–63. https://doi.org/10.1109/tpwrs.2012.2205021.
44. Lorca, Á.; Sun, X.A. Adaptive Robust Optimization With Dynamic Uncertainty Sets for Multi-Period Economic Dispatch Under Significant Wind. *IEEE Trans. Power Syst.* **2015**, *30*, 1702–1713. https://doi.org/10.1109/tpwrs.2014.2357714.
45. Zeng, B.; Zhao, L. Solving Two-Stage Robust Optimization Problems Using a Column-and-Constraint Generation Method. *Oper. Res. Lett.* **2013**, *41*, 457–461. https://doi.org/10.1016/j.orl.2013.05.003.
46. Zhang, Y.; Wang, J.; Chen, B. Chance-Constrained Two-Stage Unit Commitment Under Uncertain Load and Wind Power Output Using Bilinear Benders Decomposition. *IEEE Trans. Power Syst.* **2016**, *32*, 3637–3647. https://doi.org/10.1109/tpwrs.2017.2655078.
47. Mohajerin Esfahani, P.; Kuhn, D. Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations. *Math. Program.* **2018**, *171*, 115–166. https://doi.org/10.1007/s10107-017-1172-1.
48. Bertsimas, D.; Gupta, V.; Kallus, N. Data-Driven Robust Optimization. *Math. Program.* **2018**, *167*, 235–292. https://doi.org/10.1007/s10107-017-1125-8.
49. Bertsimas, D.; Sim, M.; Zhang, M. Adaptive Distributionally Robust Optimization. *Manag. Sci.* **2019**, *65*, 604–618. https://doi.org/10.1287/mnsc.2017.2952.
50. Yang, Y.; Wu, W. A Distributionally Robust Optimization Model for Real-Time Power Dispatch in Distribution Networks. *IEEE Trans. Smart Grid* **2019**, *10*, 3743–3752. https://doi.org/10.1109/TSG.2018.2834564.
51. Han, J.; Han, K.; Han, T.; Wang, Y.; Han, Y.; Lin, J. Data-Driven Distributionally Robust Optimization of Low-Carbon Data Center Energy Systems Considering Multi-Task Response and Renewable Energy Uncertainty. *J. Build. Eng.* **2025**, *102*, 111937. https://doi.org/10.1016/j.jobe.2025.111937.
52. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: New York, NY, USA, 2005. https://doi.org/10.1007/b106715.
53. Shafer, G.; Vovk, V. A Tutorial on Conformal Prediction. *J. Mach. Learn. Res.* **2008**, *9*, 371–421.
54. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. https://doi.org/10.1561/2200000101.
55. Stankevičiūtė, K.; Alaa, A.M.; van der Schaar, M. Conformal Time-Series Forecasting. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–14 December 2021; Volume 34, pp. 6216–6228.
56. Gibbs, I.; Candès, E.J. Adaptive Conformal Inference Under Distribution Shift. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–14 December 2021; Volume 34, pp. 1660–1672.
57. Zimmerman, R.D.; Murillo-Sánchez, C.E.; Thomas, R.J. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. https://doi.org/10.1109/TPWRS.2010.2051168.
58. Bautista-Gomez, L.; Benoit, A.; Di, S.; Herault, T.; Robert, Y.; Sun, H. A Survey on Checkpointing Strategies: Should We Always Checkpoint à la Young/Daly? *Future Gener. Comput. Syst.* **2024**, *161*, 315–328. https://doi.org/10.1016/j.future.2024.07.022.
59. Zakeri, B.; Syri, S. Electrical Energy Storage Systems: A Comparative Life Cycle Cost Analysis. *Renew. Sustain. Energy Rev.* **2015**, *42*, 569–596. https://doi.org/10.1016/j.rser.2014.10.011. Correction in *Renew. Sustain. Energy Rev.* **2016**, *53*, 1634–1635.

60. Wiese, F.; Schlecht, I.; Bunke, W.D.; Gerbaulet, C.; Hirth, L.; Jahn, M.; Kunz, F.; Lorenz, C.; Mühlenpfordt, J.; Reimann, J.; et al. Open Power System Data: Frictionless Data for Electricity System Modelling. *Appl. Energy* **2019**, *236*, 401–409. https://doi.org/10.1016/j.apenergy.2018.11.097.
61. RTE—Réseau de Transport d'Electricité. éCO2mix: Données temps réel du système électrique français. In *Open Data Réseaux Énergies (ODRÉ)*; RTE—Réseau de Transport d'Electricité: Puteaux, France, 2024.
62. Bundesnetzagentur. SMARD: Strommarktdaten der Bundesnetzagentur. In *SMARD—Strommarkt-Informationen der Bundesnetzagentur*; Bundesnetzagentur: Bonn, Germany, 2024.
63. IPCC. *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*; IPCC: Geneva, Switzerland, 2019.
64. European Environment Agency. *EMEP/EEA Air Pollutant Emission Inventory Guidebook 2019*; Technical Report 13/2019; European Environment Agency: Copenhagen, Denmark, 2019.
65. International Carbon Action Partnership; EU Emissions Trading System (EU ETS). ICAP ETS Map, 2026. 2025 average auction price EUR 73.43/tCO_2. Available online: https://icapcarbonaction.com/en/ets/eu-emissions-trading-system-eu-ets (accessed on 29 August 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.