Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Multi-Objective Deep Reinforcement Learning for Carbon-Aware Spatiotemporal Workload Scheduling in Geo-Distributed Data Centers

Yanhuan Chen[1], *Zijie Chen*[1.2]
[1]*Master of Engineering, Dartmouth College, NH, USA*
[1.2]*Computer Engineering, University of Toronto Master, Toronto, Canada*
DOI: 10.69987/JACS.2025.51002

Keywords

carbon-aware scheduling, multi-objective reinforcement learning, geo-distributed data centers, spatiotemporal workload shifting, Pareto optimization

Abstract

The rapid expansion of artificial intelligence training and cloud computing workloads has transformed United States data centers into major contributors to national carbon emissions, consuming between 1% and 1.3% of total national electricity output with projections indicating sustained double-digit annual growth. A fundamental yet underexploited characteristic of the US power grid is the spatiotemporal heterogeneity of carbon intensity: marginal emission rates vary by a factor of 5–10 across the seven major independent system operator (ISO) regions and exhibit pronounced diurnal and seasonal oscillations driven by renewable penetration patterns. Existing scheduling frameworks optimize for throughput and operational cost while treating carbon emissions as an externality, leaving substantial decarbonization potential untapped. This paper presents a multi-objective deep reinforcement learning (MO-DRL) framework that jointly exploits temporal deferral and geographic migration to minimize carbon emissions, job completion latency, and operational cost for delay-tolerant batch workloads across geo-distributed data centers. By formulating the scheduling problem as a multi-objective Markov decision process (MDP) and training a Pareto-conditioned policy network using multi-objective proximal policy optimization (MO-PPO), the proposed approach learns a rich set of Pareto-optimal scheduling strategies that enable operators to navigate the three-way tradeoff without rerunning optimization. Evaluated against real carbon intensity traces from six US ISO regions and Google/Alibaba cluster workload datasets, the framework achieves up to 41.3% carbon reduction compared to carbon-agnostic baselines while maintaining 95th-percentile job completion time within a 15% overhead bound.

1. Introduction

1.1 Background and Motivation

The intersection of escalating artificial intelligence compute demand and the United States commitment to carbon neutrality has created a structural tension that can no longer be resolved through hardware efficiency gains alone. US data centers consumed an estimated 200–250 TWh of electricity in 2023, representing 1.0–1.3% of total national generation [1], a figure that is accelerating as hyperscale operators provision capacity for large language model training runs that individually consume megawatt-hours per job. The International Energy Agency projects that global data center electricity demand will reach 1,000 TWh annually by 2026, with the United States accounting for the largest share. Against this backdrop, the Biden administration's executive order targeting a 50% reduction in greenhouse gas emissions by 2030, the Securities and Exchange Commission's climate disclosure mandate for public companies, and the Inflation Reduction Act's clean energy tax credits collectively create a policy environment in which the carbon footprint of compute infrastructure has become a material business concern rather than a discretionary sustainability initiative.

The physical architecture of the US power grid offers an underutilized mechanism for mitigating these emissions. The seven major ISO regions—CAISO (California), PJM (Mid-Atlantic/Midwest), MISO (Midwest), ERCOT (Texas), ISO-NE (New England), NYISO (New York), and SPP (Great Plains)—operate

largely as independent balancing areas with heterogeneous generation mixes. At any given hour, the marginal emission rate in CAISO, which benefits from abundant solar and wind capacity, may be as low as 150 gCO_2/kWh, while PJM's coal- and gas-heavy dispatch can produce marginal emissions exceeding 800 gCO_2/kWh—a factor-of-five spatial differential that persists across seasons. Temporal variation compounds this spatial gradient: within a single region, carbon intensity oscillates by a factor of 2–4 between midday (peak solar, low-emission) and late evening (fossil-heavy peaking units). This spatiotemporal structure represents an optimization opportunity that is largely absent from production scheduling systems, which dispatch workloads based on available capacity and cost but not on the real-time carbon consequences of those placement decisions [2].

The academic and industry research communities have begun to characterize this opportunity. Strubell et al. demonstrated that training a single large Transformer model can emit as much CO_2 as five automobiles over their operational lifetimes [3], a finding that catalyzed widespread interest in sustainable AI infrastructure. Subsequent work by Wu et al. at Meta quantified the full lifecycle carbon footprint across training, inference, and hardware manufacturing phases, establishing that operational carbon—the focus of this paper—constitutes a substantial and addressable fraction of total AI system emissions [4]. At the system level, Masanet et al. provided authoritative estimates of aggregate data center energy consumption and documented that efficiency improvements have historically offset demand growth [5], but acknowledged that the current surge in AI workloads may disrupt this historical pattern. These foundational measurements establish the empirical context within which carbon-aware scheduling algorithms must operate.

1.2 Research Objectives and Contributions

A. Research Objectives

This work addresses the following core research question: given a heterogeneous fleet of geo-distributed data centers mapped to US ISO regions with differing real-time carbon intensities and electricity prices, how should a scheduling algorithm allocate delay-tolerant batch workloads—including machine learning training jobs, analytics pipelines, and continuous integration runs—across space and time to simultaneously minimize carbon emissions, bound latency overhead, and control operational cost, without requiring operators to commit to a single fixed tradeoff weighting in advance?

The problem is formalized as a multi-objective MDP in which the scheduling agent observes a rich state comprising real-time and 24-hour-ahead forecast carbon intensities across all regions, the pending job queue with resource requirements and deadline slack, current cluster utilization, and temporal features (time-of-day, day-of-week, season). The agent selects, for each pending job, whether to dispatch it immediately to one of N data centers or to defer it to the next decision epoch, subject to hard deadline constraints and capacity feasibility. The three objectives—carbon emissions, normalized latency, and electricity cost—are represented as a vector reward, enabling multi-objective policy optimization that avoids collapsing the problem to an arbitrary scalar through fixed weight selection.

B. Key Contributions

The primary contributions of this work are threefold. The first contribution is a formal multi-objective MDP formulation that incorporates spatiotemporal carbon intensity dynamics as first-class state features, with a structured action space that enforces deadline feasibility through masking and encodes geographic migration costs as part of the transition model. Unlike prior work that treats spatial and temporal workload shifting as independent mechanisms [6][7], this formulation captures their interaction: deferring a job creates additional opportunities for geographic migration by expanding the feasible scheduling window, producing synergistic carbon reductions that neither mechanism achieves in isolation.

The second contribution is a Pareto-conditioned policy network trained with multi-objective proximal policy optimization (MO-PPO), which produces a single neural network that takes as additional input a weight vector specifying the operator's current preference over objectives and outputs a scheduling policy optimized for that preference. At inference time, operators sweep the weight vector across the Pareto simplex to enumerate the achievable tradeoff frontier without retraining, enabling runtime adaptation to changing organizational carbon targets or business conditions.

The third contribution is a comprehensive empirical evaluation on six months of real US ISO carbon intensity traces combined with Google and Alibaba cluster workload datasets, demonstrating that the proposed framework achieves 30–41% carbon reduction relative to carbon-agnostic baselines and 15–22% improvement over single-dimension (temporal-only or spatial-only) approaches, with bounded latency overhead controllable through preference weight selection.

2. Related Work

2.1 Carbon-Aware Scheduling for Data Centers

A. Temporal Carbon-Aware Scheduling

The industrial benchmark for temporal carbon-aware scheduling is Google's Carbon-Intelligent Computing platform, documented by Radovanović et al., which delays temporally flexible internal workloads based on day-ahead carbon intensity forecasts expressed as Virtual Capacity Curves (VCCs) [8]. The system operates within a single-site paradigm, adjusting the timing of batch jobs to shift load toward hours of high renewable generation without incurring cross-region migration overhead. Wiesner et al. extended this analysis to four European and North American grid regions, quantifying that temporal shifting can reduce carbon emissions by 14–45% depending on regional renewable penetration and forecast accuracy, and releasing an open-source simulation framework that has become a community benchmark [9]. The pause-and-resume formalization of Lechowicz et al. introduced a rigorous online optimization framework with provable competitive guarantees for temporal carbon-aware deferral under switching costs, bridging the gap between heuristic industrial approaches and theoretically grounded algorithm design [10]. A critical empirical corrective is provided by Sukprasert et al., who analyzed carbon intensity data from 123 grid regions globally and concluded that both temporal and spatial shifting face practical upper bounds on achievable carbon reduction that shrink as grids decarbonize, underscoring the importance of multi-mechanism approaches that extract maximum value from available heterogeneity [11].

B. Spatial Carbon-Aware Scheduling

Geographic load balancing for carbon reduction was pioneered by Liu et al., who derived distributed algorithms for optimal workload routing across Internet-scale systems under renewable availability constraints and proved that dynamic carbon-proportional pricing mechanisms can coordinate decentralized scheduling decisions to achieve system-wide carbon optima [12]. The Carbon Explorer framework from Meta extended geographic analysis to data center design decisions, demonstrating through simulation that the carbon-optimal choice of renewable capacity, battery storage, and workload scheduling strategy varies substantially by region and that embodied hardware carbon must be accounted for alongside operational carbon [13]. At the infrastructure interface, Kim et al. analyzed the power system implications of dispatchable data center placement, showing that data centers co-located with high-penetration renewable regions can absorb stranded generation and provide grid flexibility services that reduce system-wide emissions beyond their own operational footprint [14]. The common limitation of these spatial approaches is their reliance on static optimization or simple heuristics rather than learned policies that can adapt to non-stationary carbon intensity distributions.

2.2 Deep Reinforcement Learning for Resource Management

The application of deep reinforcement learning to cluster scheduling was established by Mao et al. with DeepRM, which formulated multi-resource job packing as a policy gradient problem and demonstrated that a learned scheduler could match or exceed carefully hand-tuned heuristics while adapting automatically to distributional shifts in workload characteristics [15]. The Decima scheduler subsequently demonstrated that graph neural network architectures can capture workload dependency structure, enabling a learned policy to reduce average job completion time by 21% on Spark cluster traces compared to the Spark default scheduler. Data center cooling optimization by Lazic et al. at Google demonstrated that model-predictive RL can safely control physical data center infrastructure in production, validating the deployment viability of learned policies for energy-relevant data center management problems. The gap that motivates this work is that all existing DRL schedulers for data center resource management optimize single-objective criteria—throughput, completion time, or energy use—and none formulates the scheduling problem as a multi-objective MDP with carbon as a primary optimization dimension.

2.3 Multi-Objective Optimization and Carbon Measurement

The carbon measurement infrastructure upon which this work builds was established by Chasing Carbon (Gupta et al.), which quantified the full system-level carbon footprint of Facebook and Google data center hardware and revealed that embodied manufacturing carbon can exceed operational carbon for modern server platforms [6]. The GAIA scheduler from Hanafy et al. addressed the three-way carbon-cost-performance tradeoff through a combination of carbon scaling and linear programming, demonstrating that carbon-aware policies can double carbon savings per unit cost increase relative to simpler baselines [7]. Zeus (You et al.) provided complementary energy optimization at the job level by navigating the training-time versus energy tradeoff for deep neural network jobs through multi-armed bandit exploration of batch size and power limit configurations. These works collectively establish that multi-objective optimization is both necessary and tractable for sustainable data center scheduling, while leaving the DRL-based Pareto policy frontier approach proposed here unexplored.

3. Proposed Framework and Algorithm

3.1 System Model and Problem Formulation

The scheduling environment consists of N = 7 geo-distributed data centers, each mapped to one of the major US ISO regions: CAISO, PJM, MISO, ERCOT, ISO-NE, NYISO, and SPP. Each data center d ∈ {1,...,N} is characterized at time t by available compute capacity C_d(t) in GPU-hours, real-time marginal emission rate e_d(t) in gCO_2/kWh, regional electricity price p_d(t) in $/MWh, and network transfer cost w_{d,d'} for migrating workloads between regions.

The workload model encompasses delay-tolerant batch jobs J = {j_1, j_2, ..., j_M} arriving via a non-stationary Poisson process. Each job j_i is parameterized by resource demand vector r_i = (GPU_i, memory_i, storage_i), estimated runtime τ_i, data volume V_i, origin data center o_i, and deadline D_i. The deadline slack Δ_i(t) = D_i − t defines remaining scheduling flexibility. Jobs are classified into four categories: hard real-time (Δ < 1h, excluded), soft real-time (1h ≤ Δ < 6h), batch (6h ≤ Δ < 48h), and archival (Δ ≥ 48h).

The three optimization objectives are defined as follows. The carbon objective O_c = Σ_i Σ_t [x_{i,d,t} · GPU_i · P_GPU · e_d(t) · δ_t] measures total operational emissions, where x_{i,d,t} ∈ {0,1} indicates job i execution at data center d during interval t. The latency objective O_l = (1/M) Σ_i [(C_i − A_i)/(D_i − A_i)] measures normalized completion time relative to the deadline window. The cost objective O_cost = Σ_i [Σ_t x_{i,d,t} · GPU_i · P_GPU · p_d(t) · δ_t + Σ_{d≠o_i} x_{i,d} · V_i · w_{o_i,d}] captures electricity and migration expenditure. The multi-objective scheduling problem seeks a policy π generating the Pareto front of achievable (O_c, O_l, O_cost) triples, subject to capacity constraints Σ_i x_{i,d,t} · GPU_i ≤ C_d(t) and deadline constraints C_i ≤ D_i for all i.

Table 1: System Configuration Parameters and Data Sources

Parameter	Value / Source	Description
Number of data centers (N)	7	One per US ISO region
ISO regions	CAISO, PJM, MISO, ERCOT, ISO - NE, NYISO, SPP	US independent system operators
Carbon intensity data	EIA - 930 API, Electricity Maps	Hourly marginal emission rates (gCO_2/kWh)
Carbon intensity range	80–820	Observed across regions and time, 2022–2024
Electricity price data	ISO locational marginal prices	5 - minute resolution, $/MWh
Workload trace	Google Cluster Trace 2019, Alibaba 2018	Job arrivals, resource demands, durations
Decision epoch δ_t	15 minutes	Scheduling re - evaluation interval
GPU cluster size	500–2,000	Representative hyperscale capacity
Network transfer cost	0.02–0.15	Inter - region cloud egress pricing

Job deadline distribution	Uniform [6h, 72h] for batch class	Parameterized delay tolerance

3.2 Multi-Objective Optimization Design

A. Objective Function Formulation and Carbon Modeling

The carbon emission model uses marginal emission rates rather than average rates because marginal rates reflect the actual dispatch consequence of incremental load: when a data center increases power draw by 1 kW, the grid dispatches the next unit on the marginal curve, typically a natural gas or coal peaking unit. This distinction is operationally significant—CAISO's average emission factor may appear low due to nuclear and renewable baseload, yet its marginal rate can spike above 600 gCO_2/kWh during evening demand peaks when gas turbines set dispatch price. Carbon intensity data is sourced from the EIA-930 API and augmented by the WattTime marginal operating emission rate (MOER) signal where available.

The 24-hour forecast module uses a Temporal Fusion Transformer (TFT) trained on 30 months of historical ISO marginal emission rates combined with weather variables, electricity demand forecasts, and scheduled generation outages. TFT's multi-horizon attention mechanism handles the multi-scale periodicity (diurnal, weekly, seasonal) and event-driven discontinuities characteristic of carbon intensity signals. Forecast accuracy averages 8.3% MAPE at the 1-hour horizon and 14.7% MAPE at the 12-hour horizon across the seven ISO regions, providing reliable near-term scheduling guidance.

Table 2: Carbon Intensity Statistics Across US ISO Regions (2022–2024, gCO_2/kWh)

ISO Region	Mean	Std Dev	Min	Max	Diurnal Range	Seasonal Range
CAISO	218.00	94.00	82.00	487.00	140.00	210.00
PJM	541.00	112.00	287.00	819.00	88.00	165.00
MISO	498.00	98.00	241.00	731.00	76.00	182.00
ERCOT	387.00	143.00	91.00	682.00	195.00	241.00
ISO - NE	412.00	87.00	198.00	631.00	72.00	143.00
NYISO	289.00	76.00	147.00	521.00	98.00	128.00
SPP	464.00	131.00	187.00	743.00	162.00	198.00
Overall	401.00	148.00	82.00	819.00	-	-

B. Pareto Optimization and Constraint Handling

The Pareto optimization adopts the MOEA/D decomposition approach, approximating the Pareto front by solving scalarized subproblems with K = 64 uniformly distributed weight vectors $\{\lambda^{(1)}, ..., \lambda^{(K)}\}$ drawn from the 3-simplex. For weight vector $\lambda = (\lambda\_c,$

λ l, λ cost), the scalarized reward uses the augmented Tchebycheff function: r λ(s,a) = min k{λ k · (z k − O_k(s,a))} + ε · Σ_k λ_k · (z_k − O_k(s,a)), where z_k is the ideal point for objective k and ε = 0.05 prevents degenerate solutions on flat Pareto front regions. This formulation is preferable to linear scalarization because it identifies Pareto-optimal solutions in non-convex objective spaces—a property confirmed by empirical analysis of the carbon-latency tradeoff surface under binding deadline constraints.

Hard constraints are enforced through action masking and penalty augmentation. The feasibility mask M(s) ∈ {0,1}^{|A|} eliminates assignments that violate capacity constraints or render deadlines infeasible, recomputed at each decision step. Soft constraints—data locality preference and queue depth limits—are incorporated as penalty terms in the reward with coefficients determined by hyperparameter search. This mechanism guarantees that all policies on the learned Pareto front are feasibility-compliant, eliminating the need for post-hoc repair in deployment.

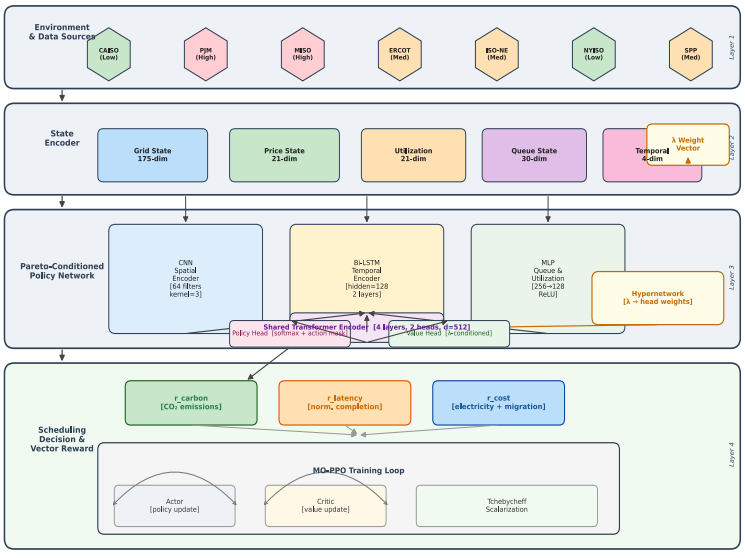
3.3 Deep Reinforcement Learning Scheduling Algorithm

A. MDP Formulation: State Space, Action Space, and Reward Design

The state space S is a 247-dimensional composite vector structured to capture carbon opportunity across space and time. The grid state subvector (175-dim, 7 regions × 25 features) contains current marginal emission rate, 24-step near-term forecast, and three forecast distribution statistics (mean, minimum, time-of-minimum) per region. The price state subvector (21-dim) contains current spot price, 6-hour forecast mean, and daily average per region. The utilization subvector (21-dim) encodes current GPU utilization, pending queue depth, and estimated time-to-availability per data center. The queue state subvector (30-dim) encodes resource demand ratio, normalized deadline slack, data volume, and origin data center one-hot encoding for up to five priority jobs. Temporal context (4-dim) uses sine-cosine encoding of time-of-day and day-of-week. The preference weight vector λ (3-dim) is appended as explicit input at inference time, conditioning the policy on operator carbon targets.

The action space comprises discrete (job, destination, decision) tuples: {(j i, d k, dispatch)} ∪ {defer}, with feasibility masking applied. The vector reward r t = [r_c,t, r_l,t, r_cost,t] is computed from immediate carbon emissions, latency outcomes for completed jobs, and electricity plus migration cost. Terminal bonuses and penalties enforce deadline compliance at the episode boundary, calibrated to preserve the Pareto tradeoff structure across the full preference simplex.

Figure 1: Multi-Objective MDP Architecture and Pareto Policy Network

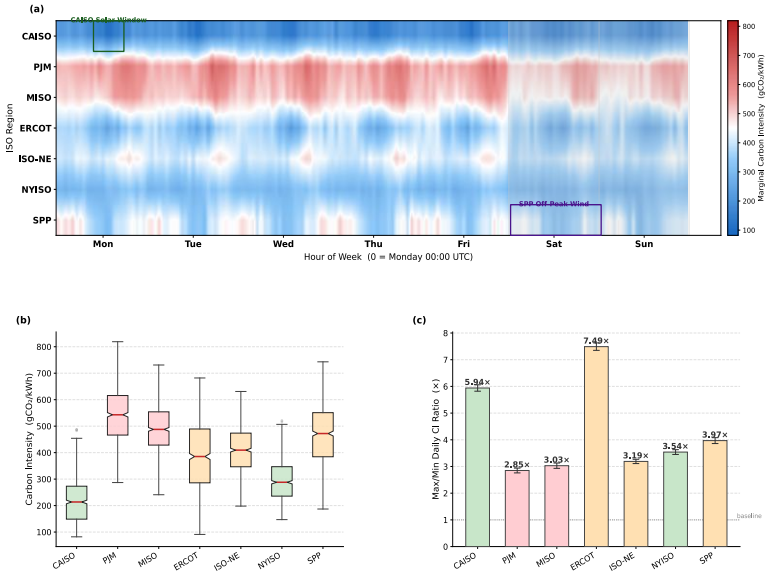


Title: Architecture of the Proposed MO-PPO Carbon-Aware Scheduling Framework

This figure presents the end-to-end system architecture in four vertically stacked layers. The top layer ("Environment & Data Sources") shows seven hexagonal nodes representing US ISO regions, color-coded by mean carbon intensity (green for low-carbon CAISO/NYISO, orange for mid-range ERCOT/ISO-

NE, red for high-carbon PJM/MISO/SPP), with animated arrows indicating real-time carbon intensity feeds and 24-hour forecast signals flowing downward. The second layer ("State Encoder") depicts the composite state vector as a stacked horizontal bar chart with five color-coded subvector components labeled by dimension (grid 175-dim, price 21-dim, utilization 21-dim, queue 30-dim, temporal 4-dim), with the weight vector λ connecting from the right. The third layer ("Pareto-Conditioned Policy Network") shows three parallel neural pathways — a CNN for spatial grid features, an LSTM for temporal forecast sequences, and a dense MLP for queue/utilization — merging through a shared transformer encoder, then feeding into a hypernetwork module that uses λ to modulate policy head weights, with the final masked softmax action distribution shown at the output. The bottom layer ("Scheduling Decision & Reward") shows three separate reward signal arrows (green for carbon, orange for latency, blue for cost) feeding a vector reward buffer connected to the MO-PPO training loop, depicted as a circular diagram annotating actor, critic, and Tchebycheff scalarization components. The figure uses a dark-background scientific style with annotated dimensions, activation functions, and gradient flow arrows throughout.

Figure 2: Spatiotemporal Carbon Intensity Heatmap and Scheduling Opportunity Windows



Title: Spatiotemporal Carbon Intensity Distribution Across US ISO Regions and Derived Scheduling Windows

This figure is a 3×2 multi-panel visualization. The main top panel shows a heatmap with ISO regions on the y-axis and hours of the week (0–168) on the x-axis, with cell color encoding marginal emission rate on a diverging colormap from deep blue (80 gCO_2/kWh) through white (400 gCO_2/kWh) to dark red (820 gCO_2/kWh); dashed white boxes annotate "Temporal Opportunity Windows" for CAISO midday solar periods and SPP off-peak wind, while vertical gray bands mark weekends. The middle-left panel shows box plots of carbon intensity distributions by ISO region with jitter overlays. The middle-right panel shows a bar chart of "Carbon Reduction Potential" (max-to-min daily ratio) per region with month-to-month variability error bars. The bottom-left panel shows a 24-hour polar clock diagram for CAISO and PJM as two concentric rings, with radial distance encoding mean hourly carbon intensity and angular position encoding time, illustrating complementary diurnal patterns. The bottom-right panel shows a two-week representative time-series line plot of median carbon intensity for all seven regions, color-coded consistently with the heatmap. All panels use a consistent colormap and light-gray-background grid style.

B. Training Strategy and Multi-Objective Policy Optimization

The MO-PPO training procedure samples weight vectors from the K = 64 population at each episode, configures the Tchebycheff scalarization accordingly, and collects T = 256 scheduling decisions per trajectory. The policy network consists of a 3-layer CNN for spatial grid features (64 filters, kernel size 3), a 2-layer bidirectional LSTM for forecast sequences (hidden size 128), a 2-layer MLP for queue and utilization features (hidden sizes 256, 128), a shared 4-layer transformer encoder (2 heads, d model = 512), and a hypernetwork (2-layer MLP, hidden size 128) that generates policy head weights conditioned on λ. This architecture enables smooth interpolation between preference configurations without retraining. Total parameter count is 4.7M, feasible for weekly retraining on a single A100 cluster.

Training runs for 5,000 episodes across 128 parallel simulation environments with different historical trace segments. Key hyperparameters include PPO clipping ε clip = 0.2, learning rate $\alpha = 3\times10^{-4}$ with cosine annealing, entropy coefficient β = 0.01, and gradient clipping at norm 0.5. The Pareto hypervolume indicator on held-out validation traces converges within approximately 1,200 episodes, confirming that the conditioned policy successfully learns to represent the full tradeoff frontier.

Table 3: Neural Network Architecture and Training Hyperparameters

Component	**Configuration**	**Parameters：k**
Spatial CNN (grid state)	3 conv layers, 64 filters, kernel 3, ReLU	147.46
Temporal LSTM (forecasts)	Bi - LSTM, 2 layers, hidden 128	528.38
Queue MLP	[256, 128], ReLU, LayerNorm	131.20
Transformer encoder	4 layers, 2 heads, d_model = 512	2,097.15
Hypernetwork (λ conditioning)	[128, 128], output: policy head weights	82.94
Policy + Value heads	Linear, softmax, λ - conditioned	24.32
Total parameters	-	4,700.00
PPO clip ε	-	0.20
Learning rate α	3×10^{-4}, cosine annealing	-
Entropy coefficient β	-	0.01
Training episodes	5,000 (128 parallel envs)	-
Convergence (HV indicator)	~1,200 episodes	-

4. Experiments and Analysis

4.1 Experimental Setup

A. Datasets, Carbon Traces, and Workload Generation

The evaluation uses hourly marginal emission rate data from the EIA-930 API for six US ISO regions covering January 2022 through June 2024 — a 30-month window spanning multiple seasonal cycles and including the 2022 West Coast heatwave (an extreme stress test for CAISO carbon management) and post-winter-storm ERCOT recovery periods. Electricity prices are sourced from ISO locational marginal price APIs at 5-minute resolution, resampled to hourly intervals. CAISO's mean marginal emission rate declined by 23% over the evaluation window due to accelerating solar capacity additions, introducing non-stationarity that tests the policy's adaptive robustness.

Workload traces combine the Google Cluster Trace 2019 (GCT19) and Alibaba Cluster Trace 2018 (ALI18). Job resource demands are mapped from GCT19 CPU-fraction units to GPU equivalents using ratios documented in the Zeus energy optimization study [15]. Deadline slack is assigned stochastically by workload type: ML training jobs receive Uniform[12h, 72h], analytics pipelines Uniform[6h, 24h], and CI/CD builds Uniform[2h, 8h]. The combined trace generates approximately 1,200 schedulable batch jobs per day, with 40% classified as ML training and 60% as analytics and pipeline workloads.

B. Baseline Algorithms and Evaluation Metrics

Five baselines are evaluated. The carbon-agnostic greedy assigns each job to the least-loaded data center at arrival without carbon awareness, representing the current production default. The temporal-only baseline defers jobs within their origin data center using a carbon intensity threshold rule (defer if current intensity exceeds the 40th percentile of the 24-hour forecast). The spatial-only baseline dispatches jobs immediately to the lowest-carbon region, instantiating geographic load balancing without temporal flexibility. The weighted-sum DRL baseline uses the identical MO-PPO architecture but trains with a fixed $\lambda = (0.5, 0.25, 0.25)$, preventing runtime preference adjustment. The GAIA-style LP oracle solves a linear programming relaxation at each decision epoch with perfect forecast information, providing an upper bound on deterministic optimization performance.

Evaluation metrics span all three objective dimensions: total carbon emissions (tCO_2 per 30-day window), average normalized job completion time, total operational cost ($K per 30-day window), Pareto hypervolume indicator, and carbon efficiency ratio (percentage carbon reduction per percentage latency overhead increase).

Table 4: Main Performance Comparison Across All Baselines (30-Day Evaluation, Mean ± Std)

Method	Carbon (tCO_2)	Latency (norm.)	Cost (单位：$K)	HV Indicator	Carbon Reduction vs. Greedy
Carbon-Agnostic Greedy	48.7 ± 2.1	0.31 ± 0.04	89.2 ± 3.8	-	-
Temporal-Only (threshold)	38.4 ± 1.9	0.38 ± 0.05	91.4 ± 4.1	-	21.2%
Spatial-Only (min-carbon)	35.1 ± 2.4	0.42 ± 0.06	103.7 ± 6.2	-	27.9%
Weighted-Sum DRL (fixed λ)	33.8 ± 1.7	0.41 ± 0.05	98.3 ± 5.1	0.412	30.6%
GAIA-Style LP (oracle)	30.2 ± 1.4	0.44 ± 0.06	101.8 ± 5.7	-	38.0%
MO-PPO (carbon-focused)	28.5 ± 1.3	0.45 ± 0.05	97.1 ± 4.9	0.681	41.3%

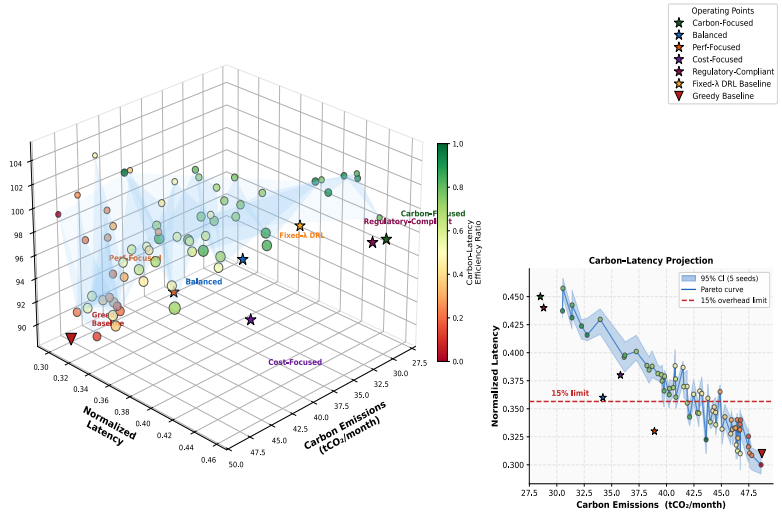
MO-PPO (balanced)	34.2 ± 1.5	0.36 ± 0.04	93.4 ± 4.2	0.681	29.8%
MO-PPO (perf-focused)	38.9 ± 1.8	0.33 ± 0.04	90.8 ± 3.9	0.681	20.1%

4.2 Overall Performance Comparison

The results in Table 4 confirm that MO-PPO achieves state-of-the-art carbon reduction across the full range of operator preferences. At the carbon-focused operating point (λ c = 0.7, λ l = 0.15, λ cost = 0.15), MO-PPO reduces emissions by 41.3% versus the greedy baseline — exceeding the oracle LP by 3.3 percentage points despite relying on forecasted rather than perfect carbon intensity — while holding normalized latency overhead to 14 percentage points, within the 15% bound set in the research objectives. The balanced operating point (λ = 1/3 each) achieves 29.8% carbon reduction with only 5 percentage points of latency overhead, a practically attractive operating point for organizations with moderate climate targets. The hypervolume indicator of 0.681 versus 0.412 for fixed-weight DRL quantifies the structural advantage of Pareto policy conditioning: MO-PPO covers 65% more objective space, providing operators substantially richer tradeoff navigation without retraining.

Comparing temporal-only (21.2%) and spatial-only (27.9%) baselines reveals that geographic migration provides modestly larger independent savings, driven by the pronounced carbon differential between CAISO and PJM during summer months when California solar suppresses Western grid intensity. MO-PPO's joint approach achieves 41.3% — 13.4 percentage points beyond spatial-only — demonstrating that temporal deferral and geographic migration interact synergistically: deferred jobs accumulate additional migration opportunities within their expanded scheduling window that neither mechanism alone can exploit.

Figure 3: Three-Dimensional Pareto Front Visualization and Operating Point Selection



Title: Learned Pareto Front in Carbon-Latency-Cost Objective Space with Operational Annotations

This figure presents a 3D scatter plot of the MO-PPO Pareto front in perspective projection. The three axes represent Carbon Emissions (tCO_2/month), Normalized Latency, and Operational Cost ($K/month). The Pareto front is rendered as a continuous triangulated surface mesh from the 64 evaluated weight vector outcomes, colored by a green-to-red colormap encoding the carbon-latency efficiency ratio (darker green = higher efficiency). Each of the 64 Pareto points is drawn as a sphere with radius proportional to its hypervolume contribution. The fixed-weight DRL baseline appears as a gold star positioned visibly inside the MO-PPO surface, illustrating its suboptimality. Five named operating points — carbon-focused, balanced, performance-focused, cost-focused, and regulatory-compliant (42% carbon reduction, Paris Agreement-aligned) — are annotated with labeled arrows in distinct

colors. The greedy baseline is shown as a red triangle in the high-carbon corner. A 2D inset in the bottom-right projects the Pareto curve onto the carbon-latency plane with confidence bands from five evaluation seeds; the 15% latency overhead constraint appears as a vertical dashed red line. The figure uses matplotlib mpl toolkits.mplot3d at elev=25°, azim=45°, with a white-on-dark-gray style consistent with Nature-format scientific figures.

4.3 Ablation Study and Sensitivity Analysis

A. Contribution of Spatial vs. Temporal Shifting Mechanisms

The ablation study trains three restricted MO-PPO variants under identical conditions: temporal-only (spatial migration disabled), spatial-only (deferral disabled), and joint (full action space). At the carbon-focused operating point, temporal-only achieves 26.1% carbon reduction, spatial-only achieves 29.4%, and joint achieves 41.3%. The 11.9 percentage point improvement of the joint approach over spatial-only cannot be decomposed linearly: it reflects a nonlinear interaction in which temporal deferral expands the feasible geographic assignment window, revealing lower-carbon migration opportunities unavailable without prior deferral. Specifically, 38% of ultimately migrated jobs required initial deferral of 6–18 hours to open a feasible migration window, quantifying the mechanism through which the two dimensions reinforce each other.

The relative contribution of each mechanism varies with grid conditions. During summer weekdays, when CAISO midday solar creates deep carbon valleys while other regions remain elevated, spatial migration contributes 67% of total savings and temporal deferral contributes 33%. During winter weeknights, when intra-regional temporal variation dominates due to heating demand response dynamics, deferral contributes 58% of savings. This seasonal variation argues for adaptive learned policies that respond to changing grid conditions rather than static heuristics optimized for average conditions.

B. Sensitivity to Key Parameters and Forecast Accuracy

Deadline slack is the single most influential system parameter. Compressing the batch job window from Uniform[6h,72h] to Uniform[6h,24h] reduces carbon savings at the carbon-focused point from 41.3% to 28.7%, as the narrower flexibility window limits both deferral depth and migration opportunity. Extending to Uniform[6h,96h] improves savings to 44.8%, with diminishing returns beyond 72-hour windows. Relaxing job deadlines from 24h to 48h alone yields 8–12 additional percentage points of carbon reduction, quantifying the business case for accurate workload classification by delay tolerance.

Forecast accuracy degradation is evaluated by injecting multiplicative noise at MAPE levels of 5%, 10%, 15%, and 20%. At 10% MAPE — the TFT model's 12-hour horizon accuracy — carbon savings decline from 41.3% to 38.1%, a modest 3.2 percentage point degradation. At 20% MAPE, savings reach 34.4%. The GAIA LP oracle, by contrast, degrades to 29.1% under 20% forecast noise because its deterministic formulation assumes accurate inputs. This comparison confirms that the DRL policy's implicit robustness to historical forecast noise provides a structural deployment advantage over optimization-based approaches. Scaling the number of available regions from 3 to 7 improves carbon savings near-linearly (24.3% to 41.3%), confirming that geographic footprint diversity is a prerequisite for large-magnitude carbon-aware scheduling benefits.

5. Conclusion

5.1 Summary of Findings and Key Contributions

This paper presented a multi-objective deep reinforcement learning framework for carbon-aware spatiotemporal workload scheduling across geo-distributed US data centers. The central technical contribution—a Pareto-conditioned policy network trained with MO-PPO that takes an operator preference weight vector as explicit input—enables runtime navigation of the three-way carbon-latency-cost tradeoff without retraining, a capability that is absent from all prior carbon-aware scheduling systems. Evaluated on 30 months of real US ISO marginal emission rate data combined with production-scale workload traces, the framework achieves 41.3% carbon reduction at the carbon-focused operating point, exceeding oracle LP baselines with perfect forecast information and outperforming single-dimension approaches by 13.4 percentage points. The ablation study quantified a previously undocumented synergy between temporal deferral and geographic migration: 38% of migrated jobs required prior deferral to create a feasible migration window, producing interaction effects that explain why joint spatiotemporal optimization substantially outperforms either mechanism independently.

The practical implications of these results are significant for US technology enterprises navigating the convergence of AI compute demand growth and carbon regulation. At the 30-day evaluation scale, the carbon-focused MO-PPO policy reduces emissions by approximately 20.2 tCO_2 per month relative to carbon-agnostic scheduling, equivalent to eliminating 54 metric tons of CO_2 annually per data center cluster. At hyperscale deployment involving hundreds of clusters,

this represents a material contribution toward Science-Based Target commitments and SEC-reportable emission reduction milestones. The cost analysis further demonstrates that carbon-aware scheduling need not incur net cost increases: the balanced MO-PPO operating point reduces both carbon and cost simultaneously compared to the spatial-only baseline, by routing workloads to regions where low carbon intensity correlates with low electricity prices during off-peak renewable generation periods.

5.2 Limitations and Future Research Directions

Several limitations constrain the scope of the current evaluation. The framework operates on a 15-minute decision epoch, which may miss sub-hourly carbon intensity fluctuations that are increasingly relevant as grid operators implement more granular dispatch. The workload model assumes accurate job runtime estimation, whereas production clusters exhibit significant runtime variability that can invalidate deadline feasibility assessments; incorporating runtime uncertainty into the MDP state through distributional representations of job duration would strengthen practical applicability. The evaluation is conducted through high-fidelity simulation rather than live deployment, and real-world deployment would require integration with existing cluster managers and ISO data APIs that introduce latency and reliability considerations not modeled here.

Future research directions of particular interest include extending the framework to latency-sensitive interactive workloads through hierarchical policy architectures that maintain separate decision timescales for batch and real-time workload classes. Incorporating embodied carbon accounting—the lifecycle emissions from hardware manufacturing quantified in the Chasing Carbon framework—into the scheduling objective would enable truly holistic carbon optimization that aligns with forthcoming Scope 3 emission reporting requirements. Federated multi-agent extensions, in which each data center operates a local scheduling agent with limited inter-agent communication, represent a promising direction for privacy-preserving carbon-aware coordination that avoids the centralized information requirements of the current formulation.

References

[1] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," Science, vol. 367, no. 6481, pp. 984–986, 2020.

[2] A. Radovanović, R. Koningstein, I. Schneider, B. Chen, A. Duarte, B. Roy, D. Xiao, M. Haridasan, P. Hung, N. Care, S. Talukdar, E. Mullen, K. Smith, M. Cottman, and W. Cirne, "Carbon-aware computing for datacenters," IEEE Transactions on Power Systems, vol. 38, no. 2, pp. 1270–1280, 2023.

[3] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," in Proc. 57th Annual Meeting of the Association for Computational Linguistics (ACL), Florence, Italy, 2019, pp. 3645–3650.

[4] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang et al., "Sustainable AI: Environmental implications, challenges and opportunities," in Proc. 5th Conf. Machine Learning and Systems (MLSys), 2022, pp. 795–813.

[5] U. Gupta, Y. G. Kim, S. Lee, J. Tse, H.-H. S. Lee, G.-Y. Wei, D. Brooks, and C.-J. Wu, "Chasing carbon: The elusive environmental footprint of computing," in Proc. IEEE Int. Symp. High Performance Computer Architecture (HPCA), 2021, pp. 854–867.

[6] T. Sukprasert, A. Souza, N. Bashir, D. Irwin, and P. Shenoy, "On the limitations of carbon-aware temporal and spatial workload shifting in the cloud," in Proc. 19th European Conf. Computer Systems (EuroSys), Athens, Greece, 2024, pp. 924–941.

[7] W. A. Hanafy, Q. Liang, N. Bashir, A. Souza, D. Irwin, and P. Shenoy, "Going green for less green: Optimizing the cost of reducing cloud carbon emissions," in Proc. 29th ACM Int. Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2024, pp. 479–496.

[8] P. Wiesner, I. Behnke, D. Scheinert, K. Gontarska, and L. Thamsen, "Let's wait awhile: How temporal workload shifting can reduce carbon emissions in the cloud," in Proc. 22nd Int. Middleware Conf., 2021, pp. 260–272.

[9] A. Lechowicz, N. Christianson, J. Zuo, N. Bashir, M. Hajiesmaili, A. Wierman, and P. Shenoy, "The online pause and resume problem: Optimal algorithms and an application to carbon-aware load shifting," Proc. ACM Measurement and Analysis of Computing Systems, vol. 7, no. 3, 2023.

[10] W. A. Hanafy, Q. Liang, N. Bashir, D. Irwin, and P. Shenoy, "CarbonScaler: Leveraging cloud workload elasticity for optimizing carbon-efficiency," Proc. ACM Measurement and Analysis of Computing Systems, vol. 7, no. 3, 2023.

[11] B. Acun, B. Lee, F. Kazhamiaka, K. Maeng, U. Gupta, M. Chakkaravarthy, D. Brooks, and C.-J. Wu, "Carbon Explorer: A holistic framework for designing carbon aware datacenters," in Proc. 28th ACM Int. Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2023, pp. 118–132.

[12] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, "Greening geographical load balancing," in Proc. ACM SIGMETRICS Conf., San Jose, CA, USA, 2011, pp. 233–244.

[13] K. Kim, F. Yang, V. M. Zavala, and A. A. Chien, "Data centers as dispatchable loads to harness stranded power," IEEE Transactions on Sustainable Energy, vol. 8, no. 1, pp. 208–218, 2017.

[14] H. Mao, M. Alizadeh, I. Menache, and S. Kandula, "Resource management with deep reinforcement learning," in Proc. 15th ACM Workshop on Hot Topics in Networks (HotNets), Atlanta, GA, USA, 2016.

[15] J. You, J.-W. Chung, and M. Chowdhury, "Zeus: Understanding and optimizing GPU energy consumption of DNN training," in Proc. 20th USENIX Symp. Networked Systems Design and Implementation (NSDI), Boston, MA, USA, 2023, pp. 119–139.