CarbonCombine: Multi-parameter Carbon-aware Scheduling

Kendall Ahern
University of Colorado Boulder

Sawyer Kennell
University of Colorado Boulder

Abstract

Data centers are typically large consumers on the U.S. electricity grid, and their demand requires both renewable and non-renewable sources. However, these load centers have specific characteristics that allow daily load shifting and compute-intensity scaling to use 'cleaner' electricity sources during the day, such as renewables, and reduce the use of 'dirty' energy during periods when fossil fuel sources are more prominent. In this work, we show how data centers can prioritize reducing carbon emissions while considering electricity prices during job scheduling windows across daily and multi-day periods. We found that there are a number of policies a scheduler can apply to carbon and price vectors to maximize utilization under specific constraints like budget and time. Our policies reflect the different priorities in applied settings, where some data centers have larger price-constrained market considerations, while others have an incentive to further reduce the carbon they incur during computation. The result of our research shows both the complexity of ongoing efforts to balance these priorities and the success different load centers can achieve when they prioritize their jobs effectively. Going forward, we can further apply the concepts in our solution to unlock more accurate insights across different markets, with varying conditions and structures.

1 Introduction

Electricity demand is increasing as the economy's needs change, and with this demand is a need for more efficient use of that power alongside expanding grid considerations. Global trends such as the transition to electric economies and instability in fuel sources indicate an increasing need for better electricity distribution, more efficient electricity use, and stronger electricity price controls over the next few decades. The International Energy Agency (IEA) has projected electricity demand to increase by 3.5 percent through 2030, and reported 2024 as the single largest increase in electricity demand in recorded history at 4.3 percent [14]. This growth in electricity demand is being driven by a number of factors: rising industrial electricity use, continued uptake of electric vehicles, increased air conditioning use due to warming, and the expansion of data centers and AI [12]. These changes are not trivial; rather, they require systemic solutions that address multiple points of concern with a single approach.

The electricity grid is a network in which generation, consumption, and arbitrage coexist to enable bidirectional flows of energy. Within this network generation sources can vary by type; from renewable to nonrenewable, to utility-scale projects over 20 megawatts (MW) and residential-scale projects as small as 150 Watts [15]. The important consideration for these technologies is how energy flows from the source to the grid, and whether that source is renewable or nonrenewable, large enough in size, and able to control the electrons it produces. The body of research our team analyzed served to underline the importance of knowing these specific features, and using them to understand the amount of carbon any such region is producing at any given time - also known as carbon intensity.

Carbon intensities vary across the grid over different timescales, depending on renewable and non-renewable generation, a region's energy mix, and the intermittent nature of solar and wind. Carbon intensity is the driving metric for how clean a region's generation might be, based on the generation it produces throughout the day, or at any one point in the day [12]. Solar arrays that rely on available sunlight will produce strong generation in the middle of the day, while wind farms might produce in the early morning, evenings, or middle of the night [12]. This is in contrast to certain fossil generation, which can provide a constant fuel supply and a more consistent generation curve. Fossil fuel's ability to generate continuous output means they are not beholden to the changes renewables see in generation, but should be considered within the scope of carbon intensity. Understanding these trends is key to better utilizing renewable energy when it becomes available during the day.

Workload flexibility allows load centers to effectively tailor their jobs to scale compute resources during times when generation is 'cleaner' than at other points in a day, or over a multi-day timescale. Data centers and other consistent electricity consumers have this flexibility at their disposal when scheduling workloads [34]. Data centers can shift workloads based on periods of peak renewable generation. When tasks are concentrated during this period of the day, the electricity fed into the grid is cleaner, and therefore the carbon intensity is lower. Although this is an excellent result for clean energy utilization, it is often not a priority for companies seeking to maximize their data center utilization [34]. For this reason, there is an incentive to incorporate additional data and insights from electricity price markets.

Electricity price market data allows generation sources on a network to predict the energy that will most likely be needed the following day [12]. These prices are publicly available and typically forecast up to 24 hours in advance, also referred to as the Day-Ahead Market (DAM) [12]. These generation sources produce energy on the same grids that data centers consume energy on. This allows for both direct and indirect carbon offsetting through the renewable market and generation sources on the network. With indirect offsetting, renewable generation sources can supply electrons to the same grid the datacenter is plugged into. Direct offsetting involves the temporal or geographic flexibility of computing resources. Cloud data centers can host delay-tolerant workloads that can be shifted to times of day when the grid is consuming more renewable generation. These data centers may also be able to shift the region from which they source their resources. Shifting either of these aspects within a scheduler can result in effective direct emission offsetting. The growing demand for this offsetting and the temporal and geographic flexibility of these workloads create a prime opportunity to combine methods to improve carbon efficiency, price awareness, and compute time.

Electricity price signals are often understood to be in contrast to carbon intensity, but are important for real-world applications that prioritize both. To take advantage of this concept, we use resource elasticity. This concept is similar to traditional cloud computing auto-scaling, where servers are dynamically allocated to a cloud application [34]. The main difference, however, is that carbon scaling is entirely dependent on forecast model outputs for carbon intensities. Determining how and when to scale is a primary decision-making function for CarbonScaler. CarbonScaler is a tool for using novel forecast data to determine a best-case schedule for periods of the day when intensities are lowest [34]. However, research in carbon intensity forecasting, electricity price market analysis, and data center scheduling does not allow for convenient weight and priority setting for price and carbon.

In this work, we introduce CarbonCombine, a set of scheduling frameworks and policies that jointly optimize inputs that balance environmental impact and cost. CarbonCombine does this by taking static data from both electricity price markets and carbon intensity, and creating novel prioritized and weighted schedules from these data sources. In other projects like CarbonCast, the model uses a tiered approach: one tier provides source electricity generation forecasts with neural network models, and the second uses these forecasts to combine weather forecasts and historical data into a 96-hour window [24]. Our paper builds on this by creating a third tier that takes the previously established forecast window, combines it with the latest electricity price market data, and allocates jobs with compute resources that can test the efficiency of the parameter priorities and constraints.

2 Background/Motivation

This section provides an overview of the evolving electricity grid, electricity markets, and data center development that impact our work in carbon-aware scheduling systems.

2.1 Grid Evolution and Renewable Variability

The global electricity grid is undergoing one of the most significant transitions since it was first constructed. For most of the 20th century, grid operations followed a relatively stable model in which dispatchable fossil-fuel generators produced electricity on demand, and operators balanced supply and demand by adjusting output from large, centralized plants with predictable costs. Now, that model is changing as variable renewable energy sources, particularly wind and solar, reshape how electricity is both generated and managed. In 2024, wind generation was 53% cheaper on average than the lowest-cost fossil fuel alternatives, while solar was approximately 41% cheaper. On top of that, 91% of newly commissioned renewable projects were more cost-effective than any new fossil fuel-based generation [16]. Unlike thermal generators, however, wind and solar output depend on variable weather conditions, and the carbon intensity of electricity in some locations depends not only on the generation mix but also on how that mix fluctuates over time due to climate [38]. The result is a more dynamic grid environment in which carbon intensity, electricity prices, and renewable availability vary on hourly and, often, sub-hourly timescales. This creates the need for systems that can anticipate and respond to these fluctuating signals to optimize for both cost savings and emissions reductions.

2.2 Electricity Markets and Price Signals

Wholesale electricity markets have evolved in parallel with these changes. Most regions across the U.S. use Locational Marginal Pricing (LMP) to determine electricity prices, reflecting the cost of delivering electricity to a specific location. This metric also incorporates generation costs, transmission congestion, and network constraints in its calculation [39] [32]. Consequently, price markets are typically structured around a two-settlement system: a day-ahead market, where

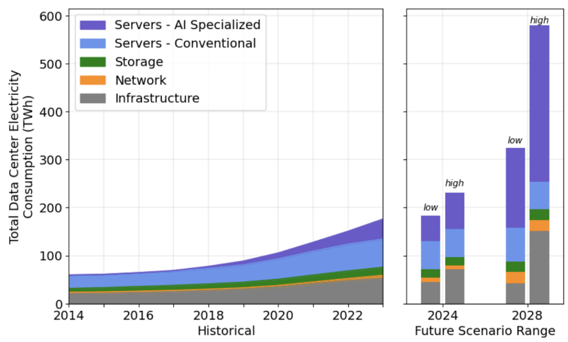

Figure 1: Total data center electricity use from 2014 to 2028.

participants commit to transactions one day in advance, and a real-time market, that adjusts for differences between expected and actual supply and demand [17]. For large electricity consumers such as data centers, aligning energy-intensive workloads with these price signals can reduce costs and, depending on grid conditions, lower associated carbon emissions.

2.3 Data Center Development and Environmental Impact

Almost simultaneously, data centers have become a major and growing source of electricity demand. In the United States, total data center electricity consumption rose from 58 TWh in 2014 to 176 TWh in 2023, and is projected to reach between 325 and 580 TWh by 2028, equivalent to roughly 6.7 to 12% of total U.S. electricity use [42].

Globally, data centers consumed an estimated 415 TWh in 2024, representing about 1.5% of total global demand, with consumption growing at approximately 12% annually over the past 5 years [13].

Data centers are highly geographically concentrated, so their growth poses a unique challenge for power systems. Within the Pennsylvania New Jersey Maryland (PJM) interconnection (one of the largest power grids in the United States), data center demand contributed to an estimated $9.3 billion increase in the 2025–2026 capacity market. This means that the projected residential electricity bill will see an increase of approximately $18 per month in western Maryland and $16 per month in Ohio [20]. At the facility level, computing equipment accounts for roughly 42% of total electricity use while cooling systems represent an additional 37% [1]. This means that even small improvements in how workloads are scheduled can lead to meaningful reductions in both energy costs and emissions [36].

The environmental impact of this growth is significant. A 2024 study of 2,132 U.S. data centers estimated their total emissions at about 105 million metric tons of CO_2, roughly 2% of national emissions, up from 31.5 million tons in 2018 [44]. Across the facilities analyzed, the average carbon intensity was 548 grams of CO_2/kWhr, approximately 50% higher than the national average across all sectors [44]. The IEA projects that data centers could account for around 1% of global CO_2 emissions by 2030, making them one of the few sectors where emissions are expected to grow [8].

2.4 Workload Flexibility and Carbon-Aware Scheduling

In response, researchers have begun exploring carbon-aware scheduling approaches that adjust computational workloads in response to changing grid conditions. The carbon intensity of electricity is not static as it varies across hours, days, and seasons. Fortunately, these patterns are closely tied to several more predictable factors, including solar generation cycles, wind availability, and demand fluctuations [24]. As forecasting methods improve, particularly through machine learning models trained on weather and historical grid data, these variations can be anticipated with higher accuracy, creating an opportunity to schedule workloads based on expected future conditions.

Approaches that optimize solely for price or solely for carbon emissions miss part of the picture, and recent work has begun to formalize this trade-off [19]. Hanafy et al. [11] investigated how to exploit temporal workload flexibility rather than relying on simple suspend-and-resume strategies. Their study shows that carbon emissions can be reduced by up to 51% without extending job completion times, while also emphasizing the importance of explicitly accounting for the cost of carbon reduction. It's important to note that these optimization strategies rely on workloads that are not strictly time-sensitive; in particular, batch workloads provide significant scheduling flexibility. Lower-priority tasks can often be shifted in time, while latency-sensitive services can be redistributed spatially across servers. In one large-scale production trace, periodic batch workloads accounted for approximately 24% of all jobs and nearly 60% of batch workloads [5]. This represents a substantial source of potential flexibility and implies that their aggregate impact on cost and emissions can be significant if optimized correctly [5] [23].

CarbonScaler was designed to take advantage of this structure by dynamically adjusting resource allocation over time rather than pausing and resuming jobs [10]. Jobs are allocated more compute capacity during low-carbon periods and less during high-carbon periods, allowing them to complete within their deadlines while still reducing total emissions. This approach translates carbon-intensity forecasts into a resource-allocation problem solved using a greedy strategy. Specifically, their experiment demonstrates benefits across machine learning and scientific workloads. Despite these advances, most existing approaches, like CarbonScaler, optimize against a single signal, typically carbon intensity, without accounting for the economic constraints that real electricity markets impose. Facilities subject to wholesale electricity prices, whether directly via market participation or indirectly through index-based contracts, incur costs that are influenced by day-ahead and real-time market structures, forward commitments, settlement prices, and location-dependent congestion effects. These factors effectively set implicit bounds on allowable energy costs, which must be taken into account when making scheduling decisions.

2.5 Limitations of Existing Approaches

Rising electricity costs and pressure to reduce emissions together introduce a co-optimization problem between carbon intensity and electricity price. These two factors are not perfectly aligned. Transmission constraints, congestion, and localized demand spikes often lead to situations in which prices diverge significantly from what carbon intensity alone would suggest.

This is the gap our work aims to address. By combining electricity price forecasts with carbon intensity forecasts and using a scheduling framework that builds on CarbonScaler's elasticity-based approach, our proposed system will enable more complete optimization. Workloads can be shifted toward periods that are both lower in carbon intensity and lower in cost, while ensuring that overall energy expenditures remain within acceptable bounds. Rather than treating cost and emissions as separate concerns, our approach will reflect the reality that data center operators must balance environmental goals with the financial constraints of electricity purchasing.

3 Design

CarbonCombine is both a price- and carbon-aware scheduling framework designed for delay-tolerant elastic workloads. The system extends the carbon-aware scheduling model introduced by CarbonScaler by incorporating electricity market signals as a first-class, high-quality scheduling input. CarbonCombine constructs schedules using a greedy optimization procedure that incrementally assigns compute resources to a series of distinct time slots.

Unlike its predecessors, which optimize solely for emissions, CarbonCombine models the interactions among carbon intensity, electricity price, workload urgency, and cost constraints. CarbonCombine enables exploration of the trade-off among these competing objectives.

3.1 System Architecture

Our design of CarbonCombine was guided by four primary goals. First, we wanted compatibility with existing carbon-aware schedulers, which meant dividing time into fixed intervals during which workloads are executed by allocating replicas to those intervals. Second, our system integrates electricity price market signals and explicitly distinguishes between day-ahead prices, which are used for planning, and real-time prices, which are used for evaluation. Third, the system supports multiple scheduling policies rather than a single objective, reflecting the relationship between carbon intensity and electricity price volatility. Fourth, our system is based on traces, so all scheduling decisions can be evaluated by replaying those historical traces. This design allows for consistent comparisons across weeks and regions while holding the workload and system parameters fixed.

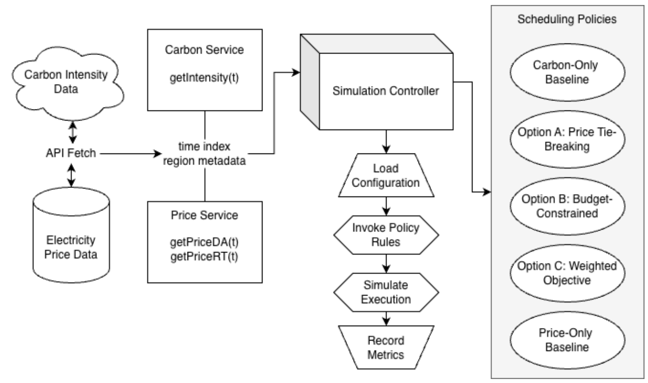

Figure 2: System architecture of Carbon-Combine.

CarbonCombine is organized as a modular pipeline that integrates time series signals, workload models, and scheduling policies. There are four primary components. First, carbon intensity and electricity price traces are loaded from external data sets via an API. Carbon traces provide time, varying emissions, and intensity, while price traces provide both day-ahead and real-time electricity prices. Second, the carbon and price services provide index access to these times, series signals, which decouple the scheduler from the underlying data representation. Third, the simulation controller orchestrates a scheduling and execution simulation by invoking a scheduling policy, simulating workload progress, and recording performance metrics. Finally, a set of policy functions computes schedules by evaluating candidate assignments under different objective functions. The system operates entirely through simulation, enabling rapid exploration across multiple weeks and regions of the data.

3.2 Scheduling Model

CarbonCombine represents a schedule as a sequence of time-indexed execution decisions that can be portrayed as a list. Each time slot t has carbon intensity C_t, a day-ahead electricity price value P_t^{DA}, and a real-time price value P_t^{RT}. Each slot is assigned a scale $s_t \in \{0,\dots,S_{\max}\}$, where $S_{\max}$ is a user-defined macro, so the schedule can be defined as $\mathbf{S} = \{(t,s_t)\}_{t=1}^{T}$. Each assignment found in this list produces work $W(t,s_t)$, energy $E(t,s_t)$, carbon $E(t,s_t)\cdot C_t$, and cost $E(t,s_t)\cdot P_t$. The goal is to build a schedule that fulfills demand, respects all deadlines, and optimizes the relevant objective

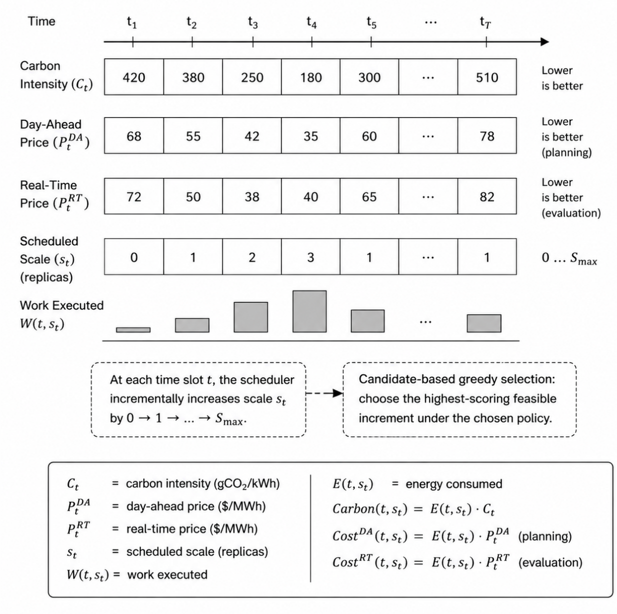

Figure 3: Policy timeline for CarbonCombine.

metrics as best as it can.

CarbonCombine employs a candidate-based greedy scheduling algorithm where each possible assignment is represented as a Candidate $= (t, s, \text{score})$. The candidates correspond to incremental increases in scale for a given time slot, and the scheduler first enumerates all feasible candidates. It then computes the score for each candidate, sorts the candidates by that score, and iteratively applies the highest-scoring feasible candidate to the slot. Then, the scheduler updates the remaining workload, and the process repeats until the workload is complete. A key constraint with CarbonCombine is that scaling is incremental, so a slot must transition from s to $s+1$ to ensure that each candidate represents only a marginal scheduling decision. This design provides two advantages: it ensures the interpretability of all scheduling decisions and enables consistent comparison across different policies.

3.3 Scheduling Polices

CarbonCombine explores the implementation of five policies that span a spectrum from carbon-only to price-only optimization.

3.3.1 CarbonScaler Baseline

This baseline policy optimizes carbon efficiency by simply maximizing work completed per unit of carbon. The score is calculated with the following equation: $\frac{\Delta W}{\Delta E \cdot C_t}$. This carbon-only policy prioritizes workload execution during low-carbon periods and serves as our environmentally friendly reference point.

3.3.2 Option A: Price Tie-Breaking

The price tie-breaking policy treats price as a secondary signal, so when two candidates have similar carbon scores, price is used as a tiebreaker. This policy preserves carbon-first behavior but opportunistically reduces costs when there's a negligible difference in carbon values. Γ is a user-defined tie-threshold and if $\frac{|s_1 - s_2|}{\max(s_1, s_2)} \leq \Gamma$, the scheduler selects the candidate with the lower price.

3.3.3 Option B: Budget-Constrained Scheduling

Option B introduces both a cost constraint and a penalty term to represent a realistic operational policy in which emissions must be reduced under explicit cost constraints. The scoring function is $\frac{(\frac{\Delta W}{\Delta C}) \cdot (1 + \gamma \cdot \frac{\Delta W}{W_{\text{remaining}}})}{1 + \beta \cdot \Delta \text{Cost}}$ and is subject to $\sum \text{Cost} \leq B$. B is the cost budget, β controls cost sensitivity, and γ controls urgency. This policy formulation combines carbon efficiency, cost awareness, and deadline urgency, and was inspired by optimization techniques learned in previous coursework.

3.3.4 Option C: Weighted Carbon-Price Objective

Option C combines carbon and price into a single scalar objective. $S_t = (\lambda \cdot \hat{C}_t) + ((1 - \lambda) \cdot \hat{P}_t)$. $\hat{C}_t)$ and $\hat{P}_t)$ are normalized carbon and price signals, respectively, and the λ value, where $\lambda \in [0, 1]$, controls the trade-off. The final score used for the scheduler becomes $\frac{W}{E \cdot S_t}$, so the policy enables continuous balancing between the two signals.

3.3.5 Price-Only Baseline

The price-only policy simply replaces carbon intensity with electricity prices. The policy schedules work in low-price periods and completely ignore emissions, serving as our economic baseline. score $= \frac{\Delta W}{\Delta E \cdot P_t^{DA}}$.

3.4 Design Decisions

A central design decision in CarbonCombine is the separation between the planning and evaluation signals. Scheduling decisions use day-ahead prices, while cost evaluation uses real-time prices. This results in a cost gap that captures the impact of forecast uncertainty and market volatility. This gap can be defined as $\Delta \text{Cost} = \text{Cost}_{RT} - \text{Cost}_{DA}$. The intention of this design is for CarbonCombine to evaluate not only cost efficiency but also the model's robustness to pricing uncertainty.

Additionally, CarbonCombine is designed to operate over multiple temporal and spatial contexts, as motivated by our background research. Multi-week replay enables evaluation across varying carbon and price conditions, while the multi-region analysis allows comparison of policy behavior under differing grid characteristics. Importantly, a scheduling policy that performs well in one region may not generalize to another region because the correlation between carbon intensity and electricity prices varies significantly across regions.

Finally, CarbonCombine separates scheduling and simulation so that evaluation remains consistent across different policies and traces. The simulation controller orchestrates the scheduling and execution. It builds a schedule using a selected policy, simulates the execution slot by slot, and tracks the cumulative metrics. The controller, however, produces a structured result that includes completion status, completion time, total energy, total carbon emissions, total cost, maximum scale, and the final schedule trace.

3.5 Design Summary

CarbonCombine builds on prior carbon-aware scheduling work by incorporating electricity and price market signals. Bringing together carbon intensity, electricity pricing, and workload urgency into a single scheduling model allows for an exploration of the trade-off between environmental impact and economic cost. CarbonCombine's core design contributions include a unified scheduling abstraction across multiple objectives, a clean separation between planning and evaluation signals, support for multiple policies, formulas, and a trace-driven evaluation methodology.

4 Evaluation

The evaluation of CarbonCombine is designed to answer three questions: How do price-aware scheduling policies change the carbon cost trade-off relative to a carbon-only baseline? How robust are these policies to the gap between day-ahead planning prices and real-time prices? Do the policies generalize across regional grid and market conditions? To answer those questions, CarbonScaler was evaluated as a carbon-only baseline, a price-only baseline was designed, and three CarbonCombine hybrid policies were designed. The experiments were conducted over multiple weekly traces. The scheduler operates over fixed time slots and assigns replica counts to each slot. Candidate assignments are scored and applied incrementally to the slots until the workload is complete. The implementation records all slot-level schedules and various aggregate metrics. The scheduler computes the carbon- and price-aware schedules, while the controller builds the schedule and replaces it, slot by slot, to record a binary completion score and various resource metrics.

4.1 Experimental Setup

All policies are evaluated on the same delay tolerant workload, which uses the same workload profile, power model, and scaling limits. The workload is modeled as an elastic batch job with a fixed amount of work to be completed. Each schedule assigns a scale value to every time slot, where the scale represents the number of replicas allocated during that slot. The controller evaluates each generated schedule by replaying the schedule over the trace. During this process, it updates total completed work, total energy, total carbon emissions, maximum scale used, active slots, and completion time. The policy is considered successful when the simulated workload completes within the available horizon.

Our final evaluation uses seven weekly traces, where each week is evaluated independently. The resulting per-week summaries are aggregated into average policy summary CSV files to analyze per-week variability and cross-week averages. CarbonCombine is evaluated on two regional cases. Ontario is the primary real data case where the evaluation uses historical Ontario carbon and electricity price traces. Price traces contain both day-ahead and real-time prices on an hourly time-scale. Texas/ERCOT is also used as a regional generalization case. The carbon and electricity price traces were obtained from a single source and converted to the same hourly format as the Ontario traces. The Texas results should be interpreted as a regional stress test for scheduler generalization, rather than as specific claims about ERCOT market outcomes, given this conversion. The evaluation included all five policy families discussed in the design section: the carbon scaler baseline, the price-only baseline, the Option A price tiebreak, the Option B budget-constrained scheduler, and the Option C weighted objective scheduler. We performed a parameter sweep across Options A, B, and C to obtain multiple results for each policy family. For Option A, we sweep the tie-threshold Γ. For Option B, we sweep the budget ratio B, the cost penalty β, and the urgency γ. For Option C, we sweep Λ.

4.2 Metrics

The CarbonCombine evaluation reports the following metrics.

- $\text{Carbon}_{\text{total}} = \sum_t E_t \cdot C_t$
  Total carbon emissions are computed by multiplying the slot, energy consumption by the carbon intensity of the slot, and lower values indicate better environmental performance.

- $\text{Cost}_{DA} = \sum_t E_t \cdot P_t^{DA}$
  The day-ahead cost is computed using prices available to the scheduler at planning time, so 24 hours in advance.

- $\text{Cost}_{RT} = \sum_t E_t \cdot P_t^{RT}$
  The real-time cost is computed after the schedule is generated, using realized real-time prices.

- $\Delta\text{Cost} = \text{Cost}_{RT} - \text{Cost}_{DA}$
  The cost app is designed to expose any market uncertainty that may be discovered. A positive gap means the real-time cost exceeded the day-ahead cost. A gap near zero means the policy was less exposed today-ahead vs. real-time price divergence, which is ideal.

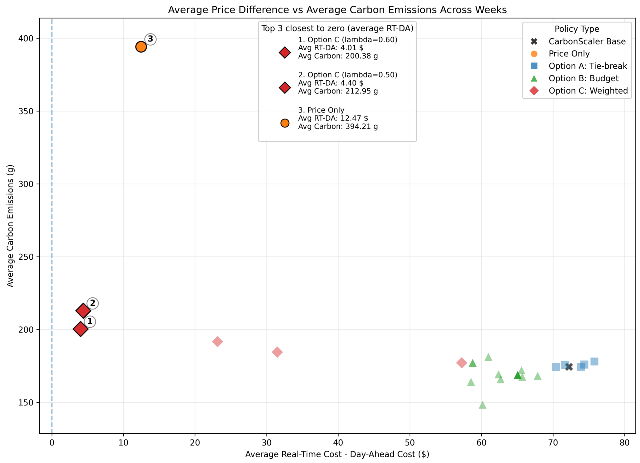

Figure 4: Average price difference and average carbon emissions across top-performing policies.

For each policy, the total carbon emissions, the day-ahead, and the real-time cost are compared against the CarbonScaler baseline. Positive savings indicate an improvement over the carbon-only baseline. Because some of the more aggressive budget settings may fail to complete within the allotted horizon, we report the completion rate across the various weekly traces. This is important because the policy that reduces carbon or cost by failing to complete the workload is not operationally useful.

5 Ontario Results

For the first round of evaluation, the policies were applied, and parameter sweeps were performed only on the Ontario data. The Ontario results show that each policy occupies a distinct region of the carbon-cost trade-off picture. The CarbonScaler baseline completes all weeks with an average carbon footprint of 174.35 gCO_2eq, an average day-ahead cost of 193.96, and an average real-time cost of 266.19. This results in a cost gap of 72.23, meaning realized real-time costs consistently exceeded the carbon-only baseline's plan.

Option A stays closest to the baseline. Across tie-break thresholds, it produces carbon emissions between roughly 174.23 and 178.06 gCO_2eq, with minimal changes in cost. This is expected because Option A only consults the price when carbon scores are nearly tied. That being said, the sweep uncovers an important anomaly that the small tie-threshold values rarely result in changes to the selected schedule. In several weeks, low thresholds produce schedules identical to the CarbonScaler baseline, and in some cases, multiple threshold values collapse to the same output. This suggests that carbon-score differences between feasible candidates often exceed the tie threshold, so the price tie-break is hardly ever used. Going forward, it would be worth reporting not only the final schedule but also how often the tie-break affects it. Additionally, it would be beneficial to explore incorporating dynamic thresholds based on the observed distribution of candidate score gaps.

Option B has the potential to deliver the strongest carbon reductions in Ontario, but it also raises the most significant feasibility concerns. Option B proves to be extremely parameter-sensitive. The most aggressive budget setting (B=0.80, $\beta = 0.01$, $\gamma = 0.50$) cuts average carbon to 148.38 gCO_2eq, more than a 14% reduction over the baseline, and lowers real-time cost by almost 19%. However, this setting only completes 14.3% of weeks. Several other low-budget Option B settings show similarly attractive carbon and cost figures, but also fail to complete in most weeks. Importantly, this led to the realization that an incomplete policy can look efficient simply because it does less work, consumes less energy, and therefore emits less carbon. Therefore, incomplete runs should not be compared directly to completed schedules unless only partial-execution outcomes are observed. Continuing on, the completion behavior of Option B reveals a sharp feasibility boundary around the $B = 0.90$ line. Budgets of 0.80 and 0.85 complete only a small fraction of weeks, while budgets at or above roughly 1.00 complete all weeks. This step-like transition suggests the budget constraint is not producing a smooth tradeoff curve as predicted. Instead, it behaves more like a phase boundary between infeasible and feasible scheduling rules. For scheduler design, this implies that budget-aware policies should include some form of feasibility prediction or budget calibration before the main scheduling pass. A future exploration into this idea would be to estimate the minimum feasible day-ahead budget needed for completion, then restrict the sweep to values above that threshold. Option B also shows parameter saturation, which points to some redundancy in our sweeping methods. Several β and γ combinations produce identical or nearly identical outputs, and some budget values become indistinguishable once the constraint is no longer binding. The scheduler operates in discrete regimes where small changes in β, γ, or the budget ratio do not affect which candidate is selected. This does not undermine the policy, but it does suggest that future sweeps should be structured around sensitivity analysis rather than uniform grids. One useful diagnostic would be a schedule uniqueness metric that counts how many distinct schedules a given sweep actually produces.

Option C shows the best cost-gap behavior in Ontario. Weighted policies with lower carbon weights, specifically λ=0.50 and λ=0.60, achieve average cost gaps of 4.40 and 4.01, respectively, far below the baseline gap of 72.23. The tradeoff, however, is higher carbon emissions. λ=0.50 emits 212.95 gCO_2eq and λ=0.60 emits 200.38 gCO_2eq which represents increases of 24.2% and 16.5% over the baseline. With this observation, Option C illustrates the tradeoff that reducing exposure to real-time price instability comes at the cost of carbon performance. The Option C sweep also reveals a plateau effect where, in some weeks, neighboring λ values produce the same schedule. This indicates that small shifts in the weighted objective are not large enough to reorder the candidate pool. Future work could once again explore adaptive or learned weighting functions that vary by week or by region rather than relying on a universal λ.

The price-only policy represents the economic extreme. It achieves the largest real-time cost savings in Ontario at 43%, with an average cost gap of 12.47, but its carbon footprint reaches 394.21 gCO_2eq, which is more than double the CarbonScaler baseline. This is one of the clearest behavioral outliers in the evaluation, and directly ties back to the motivation behind CarbonCombine's design. Cost minimization alone can produce severe carbon penalties. In some individual weeks, price-only emissions are several times higher than the carbon-only baseline, suggesting that any price-aware carbon scheduler should enforce an emissions guardrail or a hard carbon budget when price is the primary objective.

Across the Ontario results, Option B offers the best carbon savings when the budget is feasible, Option C provides the strongest reduction in cost-gap exposure, and price-only minimizes cost at unacceptable carbon expense. The central design challenge is not simply choosing a policy, but identifying parameter regions that are feasible, non-redundant, and not dominated by some alternative strategy.

5.1 Texas Results

The Texas results show that policy behavior changes significantly under different regional carbon and price conditions. The CarbonScaler baseline emits 972.17 gCO_2eq on average, compared with 174.35 gCO_2eq in Ontario. This reflects the much higher carbon intensity of the Texas case and shows that the same workload and scheduler can produce very different emissions depending on regional grid conditions.

Option A once again remains closest to the baseline. The lowest-threshold tie-break setting, $\Gamma = 0.01$, slightly improves carbon relative to the baseline, emitting 971.38 gCO_2eq, a 0.07% carbon improvement. Larger tie thresholds reduce cost more but increase carbon. For example, $\Gamma = 0.20$ reduces real-time cost by almost 17%, but increases carbon by 5%. This mirrors the Ontario behavior, where Option A has limited ability to reduce costs because it preserves carbon-first scheduling, with price incorporated only on an as-needed basis.

Option B is more cost-effective in Texas than in Ontario but still retains relatively moderate carbon behavior. Several Option B settings complete all weeks and reduce real-time cost in the range of 18–33%. For example, (B=0.90, $\beta = 0.01$, $\gamma = 0.50$) emits 1012.09 gCO_2eq and reduces real-time cost by 19%. The strongest real-time cost savings among Option B variants is achieved when B=0.90, $\beta = 0.04$, $\gamma = 0.50$, which reduces real-time cost by 33.65%, but increases carbon to 1083.31 gCO_2eq. The Texas results also reinforce the parameter-saturation behavior observed in Ontario. Option B variants often cluster tightly and some parameter values become operationally equivalent. This is useful from a systems lens because a stable cluster of Option B configurations may indicate a robust operating region. However, it also means that future work should reduce redundant sweeps and instead search for the boundaries where schedules actually change.

Option C produces the largest cost reductions among hybrid policies, but at a substantial carbon cost. With $\lambda = 0.50$, Option C reduces real-time cost by 73.25%, but carbon increases to 1570.16 gCO_2eq, a 63% increase relative to the Texas baseline. With $\lambda = 0.90$, carbon remains closer to baseline at 1046.20 gCO_2eq, but real-time cost savings fall to 26%. This confirms that Option C is highly tunable yet highly region-sensitive. The same weighted formulation that yields moderate carbon penalties in Ontario can yield much larger penalties in Texas, and vice versa.

The price-only policy once again marks the economic extreme. It reduces real-time cost by 80.77%, but carbon increases to 2044.14 gCO_2eq. This is more than twice the Texas CarbonScaler baseline and is the most severe carbon outlier in the regional evaluation. It shows that price signals alone are insufficient for environmentally responsible scheduling, especially in regions where low-cost periods do not reliably align with low-carbon periods.

The primary Takeaway from these Texas results is the understanding that the same scheduling logic can produce drastically different results when the underlying signal relationship changes. The policies designed in CarbonCombine are sensitive to the accuracy of electricity price market data and to the ratio of energy to carbon emissions on the grid.

5.2 Cross-Region Evaluation

In this section, the performance of CarbonCombine's policies is compared across the Ontario and Texas regions. The comparison highlights whether the same policies behave consistently across regions. The following figure compares average policy behavior across regions by plotting average carbon emissions against the average cost gap. Ontario points are shown as circles and Texas points as triangles, with dashed lines connecting the same policy across regions.

The most visible result is that nearly every policy shifts sharply to the right in Texas, which translates to higher carbon emissions. For example, the CarbonScaler baseline increases from 174.35 gCO_2eq in Ontario to 972.17 gCO_2eq in Texas, as discussed in the Texas Results section. This illustrates that the scheduling logic is portable, but the environmental outcome is strongly region-dependent. The second generalization result is that policies preserve their broad ordering. In both regions, price-only has the lowest cost gap but the highest carbon emissions, Option C spans a wide cost-carbon range depending on the λ, Option A remains close to the carbon-only baseline, and Option B generally occupies the middle

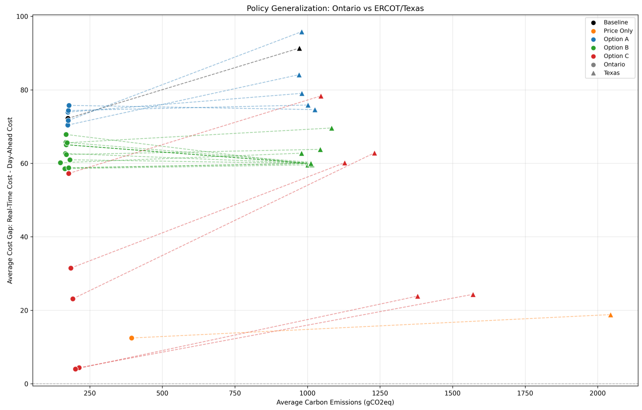

Figure 5: Policy trade-off between cost and carbon across ERCOT and Ontario regions.

Figure 6: Weekly Policy Ranking for ERCOT and Ontario Data.

of the tradeoff space. This suggests that scheduler families generalize qualitatively, even though absolute outcomes do indeed vary strongly by region.

From a quantitative perspective, Option B appears to be the most stable compromise across the regions. Its Texas points cluster around 980–1083 gCO_2eq and cost gaps around 59–70, while its Ontario points cluster around 148–181 gCO_2eq and gaps around 58–68 for most completed variants. Although the absolute carbon values vary by region, the cost-gap behavior is more stable than that of Option C or price-only. This supports the interpretation that the budget constraint acts as a regularizer, preventing the scheduler from chasing price too aggressively. Even still, Option B allows some cost reduction.

Option C is more sensitive to the region. Its low-λ variants substantially reduce the cost gap in both regions, but the carbon penalty becomes much larger in Texas. For example, λ=0.50 increases carbon by 24% in Ontario but by 63% in Texas. This suggests that the static weighted objective used in CarbonCombine may overfit the local relationship between carbon and price. A promising future direction is region-aware tuning. A λ would be learned or selected based on the historical correlation between price and carbon intensity for a specific grid.

The price-only policy is the clearest cross-region outlier. In both regions, it reduces real-time cost exposure, but in both regions, it moves far to the right in carbon. This makes it useful as a lower-bound economic reference, but not as a viable carbon-aware scheduling policy. Future designs could use price-only scheduling only as a constrained subroutine or process. One example of this could be selecting the lowest-price slots, but subjecting them to a maximum carbon threshold.

The baseline also highlights an important limitation of carbon-only scheduling. Although it remains environmentally competitive within each region, it provides no mechanism for managing the unpredictability of real-time price markets. In the regional comparison, the baseline shifts towards both higher emissions and a larger positive cost gap in the Texas case. This indicates that optimizing for carbon alone might not be sufficient when electricity price market uncertainty is a significant operational concern. Often, the ratio of renewable to nonrenewable resources on the grid has a substantial effect on the uncertainty of the price market data.

5.3 Weekly Variability

This section of analysis focuses on the weekly results which reveal meaningful temporal variation that was not shown by the average behavior calculations. Ontario weekly outcomes are concentrated at lower carbon values, generally below 500 gCO_2eq, while Texas weekly points are spread across a much wider range, often exceeding 1000 gCO_2eq. This reinforces the cross-region finding that regional carbon intensity is the dominating factor in the emissions outcomes.

The matched weekly policy-shift plot, which connects the same policy and week across Ontario and Texas, shows that regional shifts are not uniform. Some policies move primarily horizontally, indicating a carbon increase with little change in cost-gap behavior. Others shift diagonally, indicating that both carbon emissions and price uncertainty change across regions. This pattern is especially pronounced for Option C and the price-only policy.

Individual weekly plots also illustrate that no single week fully represents system-wide behavior. In week 02, many Ontario policies begin with high positive cost gaps, while Texas variants shift to much higher carbon values, with greater variation in cost-gap outcomes. In week 05, several Texas policies produce higher carbon but only moderately higher cost gaps, while Ontario points cluster at lower carbon values with a tighter spread. These week-level differences justify reporting multi-week averages rather than relying on a single representative trace. The weekly data also highlights some step-like changes in behavior. In some weeks, multiple parameter settings within the same policy family produce exactly the same schedule. This pattern is especially visible for Option B, where changes to β, γ, or even the budget sometimes yield identical outputs. This happens because the scheduler is not optimizing over a continuous solution space, but rather selecting from a finite set of candidates. If the parameter changes do not reorder those candidates, the final schedule does not change. This is an important consideration for future benchmark designs, as parameter sweeps should report not just metric differences but also how many distinct schedules were actually produced.

Finally, the weekly results highlight that apparent average improvements can depend on whether incomplete runs are included. Low-budget Option B settings often show low carbon and low cost precisely because they fail to complete. As mentioned before, this is because doing less work naturally means consuming less energy and emitting less carbon. This motivates two of the improvements suggested for future evaluations. First, plots should either filter to include only completed runs or visually distinguish incomplete runs. Second, any aggregated tables should report completion rate alongside carbon and cost metrics. Without this, a partially executed policy can appear to dominate a fully completed one, even though it is operationally infeasible.

5.4 Summary of Findings

The evaluation of CarbonCombine supports five main findings. First, as price awareness reduces cost exposure, it can increase carbon emissions depending on the policy. The price-only and lower-λ Option C policies substantially reduced the cost gap but led to a large increase in carbon emissions in both regions. This finding confirms the need for a multi-objective scheduler rather than just a price-only or carbon-only optimization technique. Second, Option A preserves carbon behavior, but provides a limited cost benefit. Because Option A only uses price as a tiebreaker, the results remain close to the CarbonScaler baseline. Additionally, the tiebreak threshold is often inactive at low Γ values, meaning that small thresholds do not meaningfully affect scheduling behavior. Third, Option B provides the most balanced trade-off, but only when feasible. The budget-constraint scheduler often reduces cost while keeping carbon closer to the baseline than the price-only or lower-λ Option C runs. However, the overly strict budget settings often failed to complete. Option B, therefore, requires an explicit feasibility analysis and should be evaluated only when using completion-aware metrics. Fourth, the response of the carbon compliance scheduler is discrete and not smoothly parameterized like initially predicted. Several sweeps contained flat regions where changing β, Γ, B, γ, or λ produced identical schedules. This suggests that CarbonCombine operates within discrete scheduling regimes determined by a finite set of candidates from the scheduling list. Future work should include a schedule-uniqueness metric and adaptive parameter searches rather than relying solely on uniform sweeps. Finally, regional generalization is not absolute, but rather qualitative. The relative ordering of policy families is similar across the Ontario and Texas regions, but the absolute emissions and the cost gaps differ substantially. This implies that carbon combine policies generalize structurally, but the specific perimeter tuning mechanism should be more region-aware. Overall, these findings show that carbon combine is indeed useful, not only as a scheduler, but also as an experimental framework for diagnosing the trade-off between carbon and price.

6 Related Work

Research done on behalf of carbon aware scheduling, reducing carbon in the scope of datacenter workloads, and shifting and optimizing job schedules are among primary related works necessary for Carbon-Combine to iterate on. The corresponding subject areas are grid carbon intensities and forecasting, weather driving and demand forecasting, and electricity and price market forecasting. These topics contain pieces of Carbon-Combine's purpose and underlying problem statement. They attempt to address any one of the problems facing data center carbon-aware scheduling, and are central to the the ability Carbon-Combine has it bringing these elements together. The two central works to our project are CarbonCast and CarbonScaler. Both projects were made by researchers at the University of Massachusetts Amherst under one lead researcher, so it's no surprise that these tools can be combined. CarbonCast uses a machine learning model that can predict electricity generation from an electricity source. Their solution uses a CNN-LSTM hybrid approach that combines weather forecasts with historical carbon intensity data to produce electricity production forecasts and carbon intensity profiles up to 96 hours into the future [25]. CarbonScaler evaluates batch workloads, but does so not by delaying compute time, but by dynamically scaling to use compute resources during periods of low carbon intensity. It can take carbon intensity forecasts like what is provided by CarbonCast and apply a scaling algorithm to scale or suspend applications in distributed batch workloads [10].

6.1 Grid Carbon-Intensity and Power-System Forecasting

Forecasting grid carbon intensity has become a critical capability for enabling carbon-aware decision making in energy systems and computing infrastructure. Research has explored alternative deep learning architectures for carbon prediction. A hybrid model that integrates a Temporal Convolutional Network with a Variational Autoencoder was proposed to capture long-range temporal dependencies and uncertainty in carbon emissions forecasting [46]. Also, a deep learning framework that combines feature selection (using Spearman correlation) with an improved denoising autoencoder was also developed. The purpose was to enhance emissions prediction while also analyzing the influence of renewable integration on emissions and electricity prices [29]. Tools like CarbonExplorer use novel methods to simulate geographic load shifting, renewable availability, workload scheduling, and infrastructure placement, but they do poorly for producing weather related data evaluation that considers a model and connects it to production [2]. More recent work has focused on improving scalability and global applicability of carbon forecasting models. CarbonX extends prior approaches by applying Time Series Foundation Models (TSFMs) to enable zero-shot forecasting of carbon intensity across more than 200 electricity grids worldwide [28]. This line of work builds on emerging TSFM research demonstrating that large pretrained time-series models can generalize across heterogeneous energy systems without extensive feature engineering [45]. Rather than relying on detailed electricity mix data, the system learns generalized temporal representations that can capture diverse grid dynamics to help enable forecasting even in regions where data might be sparse.

6.2 Weather-Driven Renewable and Demand Forecasting

Accurate modeling of both electricity demand and renewable energy variability, both of which rely heavily on weather, plays an important supporting role in forecasting grid behavior and carbon intensity. There are machine learning techniques, specifically gradient boosting regression, for hourly electricity demand forecasting using temporal and weather-based features, that outperform traditional statistical-based methods. [35]. Similarly, a hybrid architecture combining RNNs with LSTM networks has been combined to jointly model electricity load patterns and weather variables such as temperature and humidity [7].

6.3 Electricity Market and Price Forecasting

There is a parallel line of research that focuses on electricity markets, carbon-aware workload scheduling, and the economic dynamics of energy systems. CARBONDIS is a carbon-aware scheduling framework for heterogeneous GPU clusters that dynamically assigns inference workloads to different hardware resources based on real-time grid carbon intensity while adhering to performance constraints [30]. This concept has been extended by incorporating uncertainty quantification into workload shifting decisions through the UQ-Advice algorithm, which balances machine-learned forecasts with alternative fallback strategies when prediction confidence is low [18]. Other work investigates forecasting and optimization within electricity markets themselves. Machine learning approaches for day-ahead electricity price prediction demonstrate that incorporating cross-border market information and fuel prices significantly improves forecasting performance [41]. At a broader market level, a stochastic optimization framework has been developed for trading in coupled electricity and carbon markets under uncertainty in renewable generation and carbon pricing [43]. The model uses a two-stage optimization algorithm to balance profitability with emissions compliance. Similarly, Liang et al. examine integrated multi-energy systems by proposing a day-ahead scheduling framework that jointly considers uncertainties in electricity, carbon, and natural gas prices using Copula-based scenario generation [21].

Most existing systems treat carbon intensity forecasting and electricity price modeling as largely separate problems. This leaves an opportunity for a unified framework that jointly considers environmental and economic signals when optimizing energy-aware operations.

7 Conclusion

The work in CarbonCombine illustrates the tradeoffs between carbon- and price-aware scheduling. Across the regions we pulled data from, it became clear that the variability in scheduling paradigms causes for volatility in static priorities for price and carbon, and that balancing these two vectors will be an ongoing problem. Prior work in CarbonCast and CarbonScaler shows the importance of accurate data to create decisions off, and how introducing constraints on top of other work imposes real-word variables that need additional analysis to discern.

Our attempt to address these applied challenges, often made or considered by data centers across the U.S., show that jointly optimized data for carbon intensity, electricity prices, and workload constraints can be unified under different scheduling models. In other research it was often important to consider carbon as a first-class parameter among decision making, but is that logic did not seem to reflect the decisions being made in conversations around load centers as much as price. With a first-class input for electricity price signals that considers historical day-ahead data, alongside real time, CarbonCombine supports a number of scheduling policies from carbon-only to price-only optimization. The algorithm uses greedy, candidate-based selections and trace-driven simulations to explore the tradeoffs not only between policies, but parameters and regions.

Our results showed that the relationship between carbon intensity and electricity price, region to region, highly impacts the performance of any one policy. While introducing predictive price considerations, real-time volatility decreased at the cost of carbon-efficiency. In introducing policies that seek to balance parameters better, such as weighted objects and budget-constraints, the tradeoffs were made clear and present. Other policies like price tie-breaking showed little benefit, and scheduling behavior was more important around candidate ordering than parameter adjustments.

CarbonCombine showcases a series of experimental policies that reflect the complexity of real world decision making. In moving forward, modern energy systems stand to benefit from multi-objective and parameter-aware scheduling systems that better balance to goals of future load centers. With higher quality data, and better predictions, load, efficiency, and carbon, can all be considered in making decisions going forward.

8 Availability

The code for the project can be found at this link. Most of our final implementation is under the folder called `mcs_implementation`.

9 Appendix

This section details the contribution of each member of the research team. We'll walk through the project from beginning stages, to the final implementation, to showcase how our design evolved over time and resulted in the solution we present in the paper.

9.1 Kendall's Contribution

As the project began, Kendall took an accelerated approach in acquiring more information on the papers targeted in our related works phase. It became clear that the papers Kendall selected offered a lot of opportunity to be built on top of, which allowed us to move forward with a code base quickly. The project shifted when our original targeted research paper (CarbonCast) didn't have enough substance to iterate on top of. Kendall moved forward with trying other code bases, and put in an incredible amount of time testing and tweaking published code to get it operational. Kendall met with authors of the paper, and wrote detailed notes on the architecture to fill the team in on the solution. She authored much of the policy logic in the code base, and took to pulling in more data as API keys were obtained. She also helped map and diagram the architecture, and collaborated with team member to fill in gaps of understanding. In short, Kendall made a huge contribution to the design and evaluation built for CarbonCombine, and lead much of technical implementation that the team showcased.

9.2 Sawyer's Contribution

As the project began, Sawyer took time to evaluate code bases for ease of implementation and extensibility. Sawyer gathered related works and offered much of the impetus for ideas around design and solutions for code that the team evaluated. Sawyer also followed along with re-architecting code bases to evaluate as a team, but moved forward with design and documentation as Kendall accelerated towards testing code. As the code bases became more reliable, Sawyer helped support the running and testing of code, ultimately recording where the team made progress and helping channel those steps into presentations and design decisions. Sawyer helped test the final implementation, but was largely responsible for the writing and communication around our solution. In short, Sawyer made a huge contribution to the design and documentation of the code and it's output, as well as securing more data resources (like API keys and third party software), that helped refine our final solution and present it to others.

References

[1] ABI Research. *Data Center Energy Consumption Forecast, 2024–2030*. ABI Research. Mar. 2025. URL: https://www.abiresearch.com/blog/data-center-energy-consumption-forecast (visited on 03/26/2026).

[2] Bilge Acun et al. "Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters". In: *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ASPLOS 2023. Vancouver, BC, Canada: Association for Computing Machinery, 2023, pp. 118–132. ISBN: 9781450399166. DOI: 10.1145/3575693.3575754. URL: https://doi.org/10.1145/3575693.3575754.

[3] Thomas Anderson et al. "Treehouse: A Case For Carbon-Aware Datacenter Software". In: *arXiv* 3.3 (Oct. 2023), pp. 64–70. DOI: 10.1145/3630614.3630626. URL: https://doi.org/10.1145/3630614.3630626.

[4] Vahid Aryai and Mark Goldsworthy. "Day ahead carbon emission forecasting of the regional National Electricity Market using machine learning methods". In: *Eng. Appl. Artif. Intell.* 123.PB (Aug. 2023). ISSN: 0952-1976. DOI: 10.1016/j.engappai.2023.106314. URL: https://doi.org/10.1016/j.engappai.2023.106314.

[5] Yujie Cao et al. "Data-driven Flexibility Assessment for Internet Data Center towards Periodic Batch Workloads". In: *Applied Energy* 324 (2022), p. 119717. DOI: 10.1016/j.apenergy.2022.119717.

[6] Yujie Cao et al. "Data-driven flexibility assessment for internet data center towards periodic batch workloads". In: *Applied Energy* 324 (2022), p. 119665. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2022.119665. URL: https://www.sciencedirect.com/science/article/pii/S0306261922009631.

[7] Xinfang Chen et al. "Short-Term Load Forecasting and Associated Weather Variables Prediction Using ResNet-LSTM Based Deep Learning". In: *IEEE Access* (2022). DOI: 10.1109/ACCESS.2023.3236663.

[8] Josh Gabbatiss. *AI: Five Charts That Put Data-Centre Energy Use – and Emissions – into Context*. Carbon Brief. Sept. 2025. URL: https://www.carbonbrief.org/ai-five-charts-that-put-data-centre-energy-use-and-emissions-into-context/ (visited on 03/26/2026).

[9] Gianluca Guidi et al. *Environmental Burden of United States Data Centers in the Artificial Intelligence Era*. 2024. DOI: 10.48550/arxiv.2411.09786. arXiv: 2411.09786 [cs.CY]. URL: https://arxiv.org/abs/2411.09786.

[10] Walid A. Hanafy et al. "CarbonScaler: Leveraging Cloud Workload Elasticity for Optimizing Carbon-Efficiency". In: *Proc. ACM Meas. Anal. Comput. Syst.* 7.3 (Dec. 2023). DOI: 10.1145/3626788. URL: https://doi.org/10.1145/3626788.

[11] Walid A. Hanafy et al. "Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions". In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*. ASPLOS '24. New York, NY, USA: ACM, 2024. DOI: 10.1145/3620666.3651374.

[12] International Energy Agency. *Electricity 2026*. Tech. rep. Licence: CC BY 4.0. Paris: IEA, 2026. URL: https://www.iea.org/reports/electricity-2026.

[13] International Energy Agency. *Energy and AI*. Tech. rep. Licence: CC BY 4.0. Paris: IEA, 2025. URL: https://www.iea.org/reports/energy-and-ai.

[14] International Energy Agency. *Global electricity demand is set to grow strongly to 2030, underscoring need for investments in grids and flexibility*. 2025. URL: https://www.iea.org/news/global-electricity-demand-is-set-to-grow-strongly-to-2030-underscoring-need-for-investments-in-grids-and-flexibility.

[15] International Energy Agency. *Global Energy Review 2025*. Tech. rep. Licence: CC BY 4.0. Paris: IEA, 2025. URL: https://www.iea.org/reports/global-energy-review-2025.

[16] International Renewable Energy Agency. *Renewable Power Generation Costs in 2024*. Tech. rep. Abu Dhabi: IRENA, 2025. URL: https://www.irena.org/Publications/2025/Jun/Renewable-power-generation-costs-in-2024.

[17] ISO New England. *Day-Ahead and Real-Time Energy Markets*. ISO New England Inc. 2026. URL: https://www.iso-ne.com/markets-operations/markets/da-rt-energy-markets (visited on 03/26/2026).

[18] Ezra Johnson, Adam Lechowicz, and Mohammad Hajiesmaili. "Signal-Aware Workload Shifting Algorithms with Uncertainty-Quantified Predictors". In: *arXiv* (2025). DOI: 10.48550/ARXIV.2509.26511.

[19] Roman Kohút, Martin Klaučo, and Michal Kvasnica. "Unified carbon emissions and market prices forecasts of the power grid". In: *Applied Energy* 377 (2025), p. 124527. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2024.124527. URL: https://www.sciencedirect.com/science/article/pii/S030626192401910X.

[20] Rebecca Leppert. *What We Know About Energy Use at U.S. Data Centers Amid the AI Boom*. Pew Research Center. Oct. 2025. URL: https://www.pewresearch.org/short-reads/2025/10/24/what-we-know-about-energy-use-at-us-data-centers-amid-the-ai-boom/ (visited on 03/26/2026).

[21] Hongle Liang et al. "Day-ahead joint scheduling of multiple park-level integrated energy systems considering coupling uncertainty of electricity-carbon-gas prices". In: *International Journal of Electrical Power & Energy Systems* (2024). DOI: 10.1016/J.IJEPES.2024.109933.

[22] Julia Lindberg et al. "A Guide to Reducing Carbon Emissions through Data Center Geographical Load Shifting". In: *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*. e-Energy '21. Virtual Event, Italy: Association for Computing Machinery, 2021, pp. 430–436. ISBN: 9781450383332. DOI: 10.1145/3447555.3466582. URL: https://doi.org/10.1145/3447555.3466582.

[23] Wenyu Liu et al. "Online Job Scheduling Scheme for Low-Carbon Data Center Operation: An Information and Energy Nexus Perspective". In: *Applied Energy* 338 (2023), p. 120859. DOI: 10.1016/j.apenergy.2023.120859.

[24] Diptyaroop Maji, Prashant Shenoy, and Ramesh K. Sitaraman. "CarbonCast: Multi-Day Forecasting of Grid Carbon Intensity". In: *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. BuildSys '22. New York, NY, USA: ACM, 2022, pp. 198–207. DOI: 10.1145/3563357.3564079.

[25] Diptyaroop Maji, Prashant Shenoy, and Ramesh K. Sitaraman. "Multi-Day Forecasting of Electric Grid Carbon Intensity Using Machine Learning". In: *ACM SIGEnergy Energy Informatics Review* (2023). DOI: 10.1145/3607114.3607117.

[26] Diptyaroop Maji, Prashant Shenoy, and Ramesh K. Sitaraman. "Multi-Day Forecasting of Electric Grid Carbon Intensity Using Machine Learning". In: *SIGENERGY Energy Inform. Rev.* 3.2 (June 2023), pp. 19–33. DOI: 10.1145/3607114.3607117. URL: https://doi.org/10.1145/3607114.3607117.

[27] Diptyaroop Maji, Ramesh K. Sitaraman, and Prashant Shenoy. "DACF: day-ahead carbon intensity forecasting of power grids using machine learning". In: *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*. e-Energy '22. Virtual Event: Association for Computing Machinery, 2022, pp. 188–192. ISBN: 9781450393973. DOI: 10.1145/3538637.3538849. URL: https://doi.org/10.1145/3538637.3538849.

[28] Diptyaroop Maji et al. "CarbonX: An Open-Source Tool for Computational Decarbonization Using Time Series Foundation Models". In: *arXiv (Cornell University)* (2025). DOI: 10.48550/ARXIV.2510.01521.

[29] Sana Mujeeb et al. "Deep learning based carbon emissions forecasting and renewable energy's impact quantification". In: *Iet Renewable Power Generation* (2023). DOI: 10.1049/RPG2.12641.

[30] Seyed Morteza Nabavinejad et al. "CarbonDIS: Carbon-Aware DNN Inference Scheduling on Heterogeneous GPUs". In: *IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing* (2025). DOI: 10.1109/CCGRID64434.2025.00032.

[31] Michał Narajewski and Florian Ziel. "Econometric modelling and forecasting of intraday electricity prices". In: *Journal of Commodity Markets* 19 (Sept. 2020), p. 100107. ISSN: 24058513. DOI: 10.1016/j.jcomm.2019.100107. URL: https://linkinghub.elsevier.com/retrieve/pii/S2405851319300728 (visited on 03/03/2026).

[32] Andrew L. Ott. "Experience with PJM Market Operation, System Design, and Implementation". In: *IEEE Transactions on Power Systems* 18.2 (2003), pp. 528–534. DOI: 10.1109/TPWRS.2003.810692.

[33] Xiaoyang Qu et al. " GreenMatch: Renewable-Aware Workload Scheduling for Massive Storage Systems ". In: *2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. Los Alamitos, CA, USA: IEEE Computer Society, May 2016, pp. 403–412. DOI: 10.1109/IPDPS.2016.24. URL: https://doi.ieeecomputersociety.org/10.1109/IPDPS.2016.24.

[34] Rocky Mountain Institute. *Fast, Flexible Solutions for Data Centers*. July 2025. URL: https://rmi.org/fast-flexible-solutions-for-data-centers/.

[35] B. Sahin et al. "Predicting Energy Demand Using Machine Learning: Exploring Temporal and Weather-Related Patterns, Variations, and Impacts". In: *IEEE Access* (2023). DOI: 10.1109/ACCESS.2024.3370442.

[36] Arman Shehabi et al. *2024 United States Data Center Energy Usage Report*. Tech. rep. LBNL-2001637. Berkeley, CA: Lawrence Berkeley National Laboratory, 2024. URL: https://eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report_1.pdf.

[37] Abel Souza et al. "CASPER: Carbon-Aware Scheduling and Provisioning for Distributed Web Services". In: *Proceedings of the 14th International Green and Sustainable Computing Conference*. IGSC '23. Toronto, ON, Canada: Association for Computing Machinery, 2024, pp. 67–73. ISBN: 9798400716690. DOI: 10.1145/3634769.3634812. URL: https://doi.org/10.1145/3634769.3634812.

[38] Thanathorn Sukprasert et al. "On the Limitations of Carbon-Aware Temporal and Spatial Workload Shifting in the Cloud". In: *Proceedings of the Nineteenth European Conference on Computer Systems*. EuroSys '24. ACM, Apr. 2024, pp. 924–941. DOI: 10.1145/3627703.3650079. URL: http://dx.doi.org/10.1145/3627703.3650079.

[39] Leigh Tesfatsion. "Locational Marginal Pricing: A Fundamental Reconsideration". In: *IEEE Open Access Journal of Power and Energy* 11 (2024), pp. 104–116. DOI: 10.1109/OAJPE.2024.3361751.

[40] Shivani Tripathi et al. "Workload Shifting Based on Low Carbon Intensity Periods: A Framework for Reducing Carbon Emissions in Cloud Computing". In: *2023 IEEE International Conference on Big Data (BigData)*. 2023, pp. 3387–3395. DOI: 10.1109/BigData59044.2023.10386447.

[41] Léonard Tschora et al. "Electricity price forecasting on the day-ahead market using machine learning". In: *Applied Energy* 313 (May 2022), p. 118752. ISSN: 03062619. DOI: 10.1016/j.apenergy.2022.118752. URL: https://linkinghub.elsevier.com/retrieve/pii/S0306261922002057 (visited on 03/03/2026).

[42] U.S. Department of Energy. *DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers*. U.S. Department of Energy. Dec. 2024. URL: https://www.energy.gov/articles/doe-releases-new-report-evaluating-increase-electricity-demand-data-centers (visited on 03/26/2026).

[43] Jing Wang et al. "Research on Uncertainty Modeling and Optimization Trading Algorithms in the Electricity-Carbon Coupled Market". In: *2024 3rd International Conference on Clean Energy Storage and Power Engineering (CESPE)* (2024). DOI: 10.1109/CESPE64643.2024.00054.

[44] Miguel Yañez-Barnuevo. *Data Center Energy Needs Could Upend Power Grids and Threaten the Climate*. Environmental and Energy Study Institute. Apr. 2025. URL: https://www.eesi.org/articles/view/data-center-energy-needs-are-upending-power-grids-and-threatening-the-climate (visited on 03/26/2026).

[45] Lifan Zhao et al. "Less is More: Unlocking Specialization of Time Series Foundation Models via Structured Pruning". In: *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. 2025. URL: https://openreview.net/forum?id=jy4bBsr1Jc.

[46] Qian Zhao et al. "Carbon Emission Prediction in Power Systems Using A VAE-TCN Hybrid Architecture". In: *Asia-Pacific Conference on Intelligent Robot Systems* (2025). DOI: 10.1109/ACIRS66343.2025.11360854.