Optimizing Workload Migration for Carbon and Cost Reductions Under Grid Constraints

New Insights and a Practical Evaluation Framework

By Can Hankendi and Ayse K. Coskun,

Digital Object Identifier 10.1109/ESM.2025.3635119
Date of current version: 29 January 2026

AS CLOUD-SCALE ARTIFICIAL INTELLIGENCE (AI) and data processing workloads continue to surge, data center operators are increasingly exploring workload migration strategies to optimize energy costs and reduce environmental impact. While prior work has addressed cost- or carbon-aware migration independently, fewer studies have incorporated power grid strain as an operational constraint alongside these objectives. This article introduces a multiobjective optimization framework for cross-regional workload migration that minimizes electricity cost and carbon emissions, while enforcing strict constraints on grid strain and workload performance. Our approach uses real-world data from

©SHUTTERSTOCK.COM/ROBUART

U.S. independent system operators (ISOs) to evaluate the feasibility and effectiveness of migrating workloads across regions under these combined considerations. We model migration overheads, including penalties from latency and runtime extension, and show how they influence the net benefit of shifting workloads. Our experiments show that up to 13% carbon and up to 15% cost reductions can be achieved while ensuring that no destination region exceeds safe grid operating thresholds. Furthermore, we highlight temporal and seasonal patterns that impact migration opportunities. This work provides new insights for sustainable workload migration strategies in geodistributed infrastructures and introduces a practical framework for evaluating carbon- and cost-aware migration decisions under explicit grid strain constraints.

Introduction

The rapid growth of AI workloads has imposed significant stress on data center infrastructures. The number of hyperscale and enterprise data centers is expected to increase by 63%, and the total data center electricity capacity is projected to triple by 2030 to satisfy the increasing demand (Synergy Research Group, 2025). Such growth poses significant challenges for power grids across many regions in the United States (Haslak, 2016; National Renewable Energy Laboratory, 2024). At the same time, increasing operational costs and carbon emissions have become major concerns for data center operators.

Modern data centers are geographically distributed across regions served by independent electricity markets and grids. This geodistribution creates opportunities to dynamically shift workloads among regions to exploit spatial and temporal variations in electricity cost, carbon intensity, and grid capacity. For instance, migrating workloads from a high-cost, fossil-fueled region to another region with abundant renewable generation can significantly reduce both cost and carbon footprint. Beyond these direct benefits, workload migration can also help avoid exacerbating stress on regional power grids, especially during peak demand periods or grid emergencies. However, migration decisions that fail to account for grid operating limits risk destabilizing load-to-capacity ratios or other grid stress indicators as a hard operational constraint rather than an optimization objective. This ensures that cost and carbon savings are only pursued if they do not push any region beyond safe operating thresholds. Migration decisions must also respect workload performance constraints as moving large jobs across regions can incur latency penalties or runtime extensions that negate potential benefits (New York Independent System Operator, 2025). Achieving sustainable and stable operation therefore requires a coordinated approach that accounts for economic, environmental, and grid operational factors simultaneously.

In this article, we present an optimization framework that minimizes electricity cost and carbon emissions for cross-regional workload migration while enforcing strict constraints on both grid strain[1] and workload performance, as shown in Figure 1. Using real-world electricity price and carbon intensity data from major U.S. ISOs, including CAISO, ERCOT, NYISO, PJM, MISO, SPP, and ISO-NE, we evaluate the feasibility and impact of these migrations under different penalty scenarios. Our framework incorporates a region-specific strain model that accounts for power utilization and capacity limits, both quantifying the relief provided to the source region and ensuring no violation of destination constraints. We perform hourly level analysis over multiple seasons to characterize temporal and geographic variability.

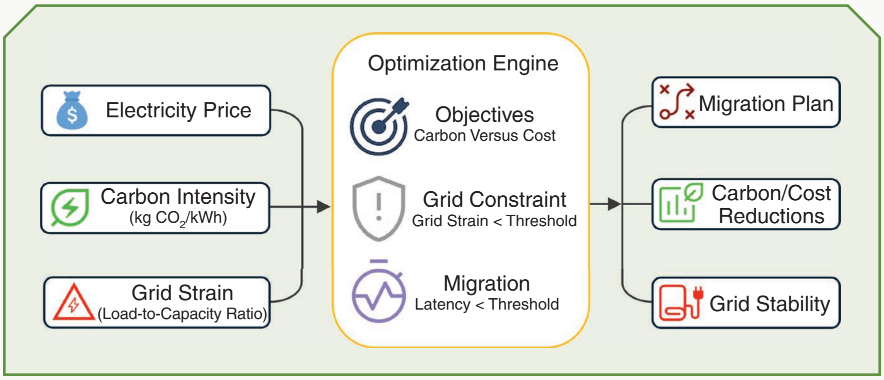


figure 1. An overview of the workload migration optimization framework. The proposed framework integrates regional electricity price, carbon intensity, and grid strain to inform carbon- and cost-aware workload migration decisions across data centers.

---

[1] Grid strain refers to the relative stress on an electricity system, measured by the ratio of load to available capacity.

Our contributions are as follows:

- We propose a workload migration optimization model that minimizes cost and carbon emissions under explicit grid strain and performance constraints.
- We present an empirical evaluation across six U.S. ISOs using real-time data to quantify achievable benefits and limitations of migration.
- We identify high-impact migration opportunities through heat maps and seasonal analyses, providing actionable insights for cloud and grid operators seeking to improve data center sustainability without compromising grid reliability.

Formulating an Optimization Model for Workload Migration With Cost and Carbon Objectives Under Strain and Performance Constraints

Optimizing modern data center operations is a challenging task, where operational costs, carbon emissions, and grid reliability are interdependent. The dynamic nature of electricity markets, combined with the variability in carbon intensity across regions, presents both challenges and opportunities for optimizing workload distribution. Our goal is to develop a model that enables data centers to make informed migration decisions that minimize electricity cost and carbon emissions, while ensuring that performance degradation and power grid strain remain within acceptable limits.

Unlike traditional load balancing approaches that primarily focus on latency and utilization, our model explicitly incorporates migration overheads and enforces grid strain and performance as hard constraints. This approach ensures that environmental and cost benefits are not achieved at the expense of operational safety or quality of service.

To capture the effect of network bandwidth on workload migration, we modeled latency as proportional to workload size and inversely proportional to bandwidth (Table 1). We first introduce a latency-aware component that accounts for workload size and available interregion bandwidth, enabling us to quantify the performance impact of migration. The migration delay depends on two factors: the size of the workload and the available bandwidth between regions. Larger workloads or lower bandwidths naturally increase the transfer time. The migration-induced latency is estimated by dividing the workload size by the available interregion bandwidth. This provides a simple way to approximate how long it would take to move workloads between regions. This latency is compared against a maximum allowable latency to ensure that migrations do not violate performance constraints. To compute the electricity cost in the destination region, we multiply the workload's energy demand with the local electricity price and add any migration energy overhead. Similarly, carbon emissions are calculated based on the workload's energy demand and the carbon intensity of the destination grid, adjusted for migration overhead.

table 1. Model parameters for migration optimization under strain and performance constraints.

Factor	Description
Workload demand (kWh)	Amount of energy required to run the workload in the source region
Electricity price (US$/kWh)	Local price of electricity in the destination region
Carbon intensity (kg CO_2/kWh)	Carbon emissions per unit of electricity in the destination grid
Grid strain index (0–1)	Measure of how close the region's grid is to its maximum capacity
Maximum strain threshold	Upper limit of allowable strain to maintain grid reliability
Migration overhead (kWh)	Additional energy used to transfer the workload across regions
Workload size (GB)	Data size of the workload that must be migrated
Available bandwidth (GB/s)	Network bandwidth available between regions
Latency threshold (s)	Maximum allowable delay during migration
Weighting factor	Adjusts how much priority is given to cost versus carbon emissions

Grid Strain Constraint

Grid strain reflects how heavily a region's electrical infrastructure is being used, typically expressed as the ratio of demand to available capacity. When this ratio is high, the grid is operating close to its limit, which increases the risk of instability or even outages. If additional computing workloads are migrated into a

> This geodistribution creates opportunities to dynamically shift workloads among regions to exploit spatial and temporal variations in electricity cost, carbon intensity, and grid capacity.

region under these conditions, they can push the system beyond safe operating levels. To prevent this, our framework evaluates the impact of each potential migration on the destination grid and only permits transfers if the resulting load remains below a defined safe threshold. This ensures that cost and carbon savings are achieved without compromising grid reliability.

Optimization Objective

The optimization minimizes a weighted combination of cost and carbon, where the relative importance of each can be adjusted using a tunable parameter. Migration is only considered feasible if it meets both the grid strain and workload performance constraints.

This formulation ensures that migration decisions deliver cost and carbon benefits without overloading destination grids or violating workload performance requirements. Strain relief can still be measured after solving the optimization, but it is no longer directly optimized, therefore making it a constraint rather than an explicit goal.

Cross-Regional Workload Migration Methodology

In this section, we present our approach for modeling and optimizing spatial workload migration across U.S. electricity markets, with the aim of optimizing electricity costs, carbon emissions, and grid strain. Our methodology utilizes real-world grid data with a formal optimization framework to make intelligent workload migration decisions. Our model leverages linear programming to identify migration strategies that minimize electricity costs and carbon emissions, while simultaneously relieving strain on overburdened grid regions. The objective is a weighted sum of cost and carbon, defined as $\min_j [\alpha \cdot \text{Carbon}_{ij} + (1-\alpha) \cdot \text{Cost}_{ij}]$. Furthermore, our model incorporates latency as a function of bandwidth and data size, ensuring that performance considerations are taken into account. By combining real-world data from multiple ISOs with a multi-objective optimization formulation, we provide a data-driven solution for carbon- and cost-aware workload migration across different regions.

Our methodology consists of the following key stages:

1) *Data collection*: Aggregating electricity price, carbon intensity, and strain metrics from public sources for all major U.S. ISOs.
2) *Normalization and preprocessing*: Standardizing data for use in optimization by aligning temporal and spatial granularity.
3) *Latency and migration penalty modeling*: Assigning latency penalties between region pairs based on geographical proximity and network performance assumptions.
4) *Optimization formulation*: Constructing a linear programming model incorporating cost, carbon, and strain objectives with migration constraints.
5) *Scenario-based evaluation*: Running simulations across various seasons, demand profiles, and penalty levels to evaluate tradeoffs and generate insights.

We collect hourly electricity market data from the following regional ISOs: CAISO (California Independent System Operator, 2025), ERCOT (Electric Reliability Council of Texas, 2025), PJM (PJM Interconnection, 2025), MISO (GridStatus.io, MISO Live Dashboard and Price Map, 2025), SPP (GridStatus.io, SPP Live Dashboard and Price Map, 2025), NY-ISO (New York Independent System Operator, 2025), and ISO-NE (ISO New England, 2025). In addition, for each ISO, we obtain the hourly carbon intensity data from publicly available dashboards (Electricity Maps, 2024) and the strain index (S_r), which is estimated using normalized load-to-capacity ratios.

All electricity price, carbon intensity, and grid strain data were standardized to an hourly format so that regions could be compared consistently. Any missing data points were filled using standard interpolation methods to ensure continuity. To estimate the time it takes to move workloads, we considered both the size of the workload and the distance between regions. Nearby regions were assumed to have short transfer delays, while distant pairs faced longer ones. We also set an upper limit on allowable delay so that performance-sensitive jobs would not be disrupted.

Our framework balances electricity cost and carbon emissions while ensuring that grid reliability and workload performance are minimally affected. For each hour in the dataset, it evaluates regional prices, carbon intensity, and grid conditions, then recommends the best placement for workloads. Operators can adjust priorities, for instance, placing more emphasis on carbon reduction versus cost savings, depending on their objectives. We evaluate our model across 1) different seasons (winter, spring, summer, and fall) to capture seasonal variability in cost and emissions and 2) ISO-specific migration paths to identify the most favorable destination that provides the highest cost and/or carbon benefits under grid strain constraints.

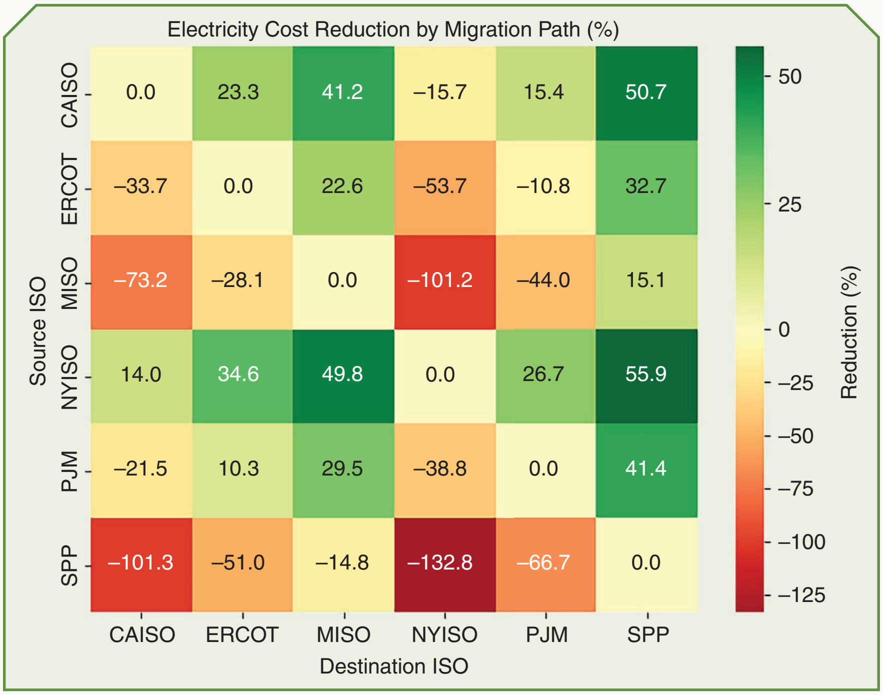
To evaluate our migration framework under realistic conditions, we model workloads that are similar to large language model training workloads. For our simulations, we evaluate a workload with dataset sizes of 2 TB, training durations of approximately 100 hours, and total energy consumption per job reaching 200 kWh, which aligns with publicly reported metrics for models such as GPT-2 (Radford et al., 2019)

Experimental Results

In this section, we present the results of our workload migration optimization framework across major U.S. ISOs. Our analysis is based on hourly electricity price and carbon intensity data collected over multiple seasons and incorporates realistic migration penalties ranging from 0% to 5%. We focus on three primary metrics, electricity cost savings, carbon emissions reduction, and grid stress violations, while observing how these metrics are affected by migration direction, latency penalties, and seasonal conditions.

Electricity Cost Savings Across ISOs

We first evaluate the cost savings achievable by migrating workloads from a source region (e.g., ERCOT) to other ISOs. Figure 2 shows the average percentage reduction in electricity cost under a 0% migration penalty across all region pairs. We observe several takeaways in the cost-saving potential of interregional workload migration. Substantial savings are observed when shifting workloads from high-cost regions such as NYISO and CAISO to areas with more stable and lower electricity prices, including MISO and SPP. These gains, however, are not uniform throughout the year, as the amount of savings depends on the season. For example, CAISO exhibits its highest cost-saving potential during the summer months, coinciding with elevated electricity prices driven by peak demand and increased grid stress. Furthermore, the advantage of migration can diminish significantly under even modest latency penalties. A performance penalty as low as 2.5% is sufficient to offset the benefits in scenarios where regional price differences are marginal, highlighting the need to account for performance constraints when making migration decisions.



Source ISO	CAISO	ERCOT	MISO	NYISO	PJM	SPP
CAISO	0.0	23.3	41.2	−15.7	15.4	50.7
ERCOT	−33.7	0.0	22.6	−53.7	−10.8	32.7
MISO	−73.2	−28.1	0.0	−101.2	−44.0	15.1
NYISO	14.0	34.6	49.8	0.0	26.7	55.9
PJM	−21.5	10.3	29.5	−38.8	0.0	41.4
SPP	−101.3	−51.0	−14.8	−132.8	−66.7	0.0

figure 2. A heat map of the maximum electricity cost savings from migrating workloads across ISO regions (0% penalty). Notably, migrating workloads from NYISO provides the most significant reductions.

Carbon Emissions Reductions

Carbon savings from migrating workloads among regions are largely influenced by temporal and spatial variations in carbon intensity, defined as the amount of CO_2 emitted per unit of electricity consumed. As shown in Figure 3, different seasons and regions present varying opportunities for reducing emissions. Some regions, such as ERCOT and SPP, frequently emerge as favorable migration targets due to their relatively low average carbon intensity, particularly in the spring and fall when renewable generation (i.e., wind) is more abundant. Conversely, regions that are heavily reliant on fossil fuels, such as MISO during winter months, tend to exhibit higher emissions per kilowatt-hour. Migrating workloads from these carbon-intensive regions to cleaner grids like CAISO or ISO-NE during periods of high renewable output results in some of the greatest observed carbon savings.

Another important observation is that unlike electricity costs, carbon reductions are less affected by small performance tradeoffs such as latency penalties. Even if a job runs a bit slower due to being migrated across regions, the carbon impact is primarily determined by where the energy is coming from and how clean that energy is. This makes carbon-aware workload migration a promising strategy even in cases where performance sensitivity might otherwise limit cost-based decisions.

Seasonal Trends in Electricity Cost Reduction

Workload migration potential is heavily influenced by seasonal fluctuations in electricity prices, which vary due to changing demand, renewable generation, and fuel costs (International Energy Agency, 2023; Radovanović et al., 2022). Figure 4 illustrates the seasonal cost reduction trends across the seven ISOs. From Figure 4, we observe that cost reduction potential is consistently highest in the summer, particularly in CAISO, reflecting elevated peak-time prices and stress on the grid due to cooling demand. For NYISO, the largest cost reductions occur in spring, when wider price spreads emerge despite lower summer peaks. Furthermore, several regions, such as ERCOT and MISO, exhibit lower benefits in spring and fall, possibly due to more stable demand and abundant renewable supply. Finally, we see that CAISO and ISO-NE show secondary peaks in winter due to heating-related demand spikes. These findings suggest that a static migration policy would miss significant savings opportunities. Instead, workload migration strategies should adjust dynamically based on seasonal market behavior.

Seasonal Trends in Carbon Emissions Reduction

Seasonal fluctuations in renewable generation and fossil fuel usage across regional grids play a critical role in determining the carbon reduction opportunities. As illustrated in Figure 5, the average emissions savings from workload migration vary significantly by ISO and season. CAISO and SPP achieve the highest reductions during summer, exceeding 23%, when solar and wind generation peak and carbon intensity decline. These findings highlight the importance of aligning migration strategies with seasonal generation profiles as carbon intensity can vary dramatically even within the same region over time. PJM also shows strong spring and summer reductions, driven by higher renewable dispatch and lower

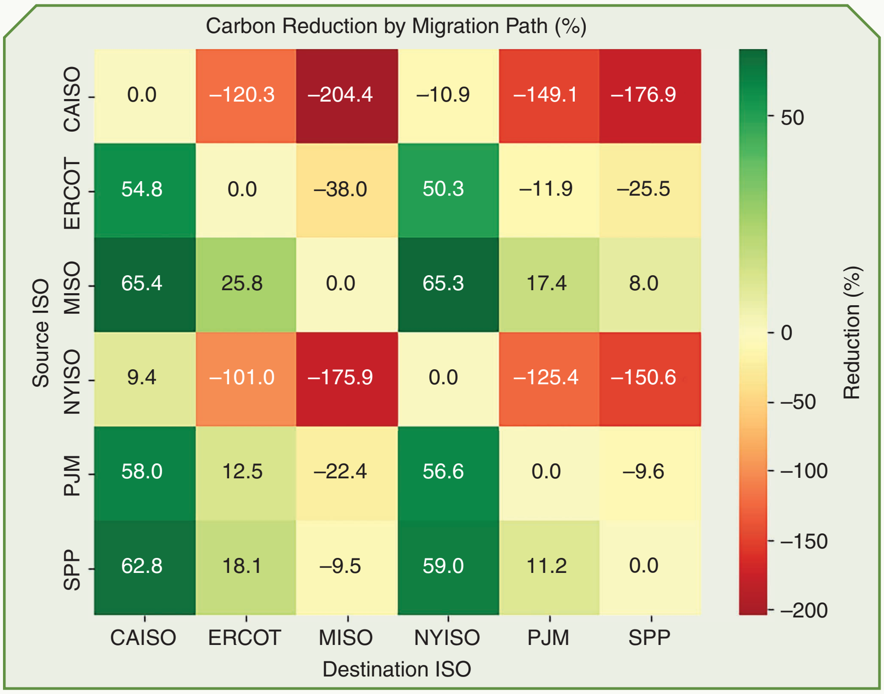


figure 3. A heat map of carbon savings by migration path (maximum over all seasons). Notably, migrations to CAISO consistently achieve positive carbon savings, driven by its cleaner energy mix.

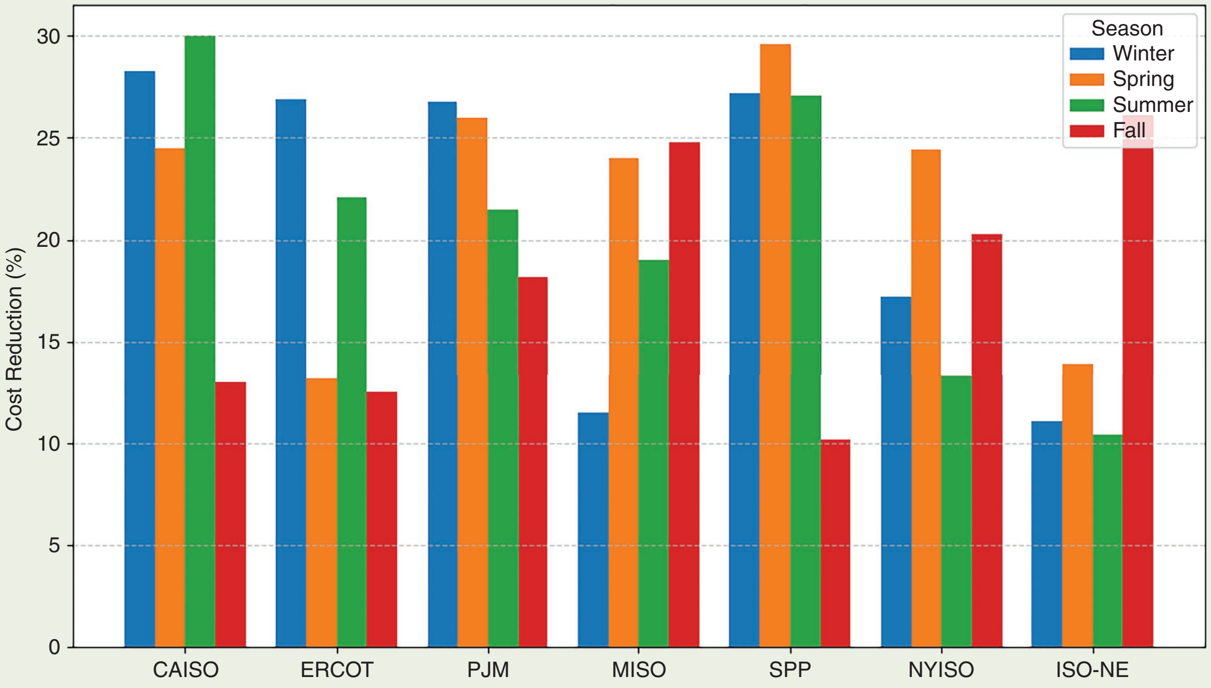


figure 4. Seasonal electricity cost reduction across ISOs. The chart highlights strong seasonal variation across ISOs, with some regions such as CAISO and SPP showing larger cost savings in the summer months when renewable generation peaks.

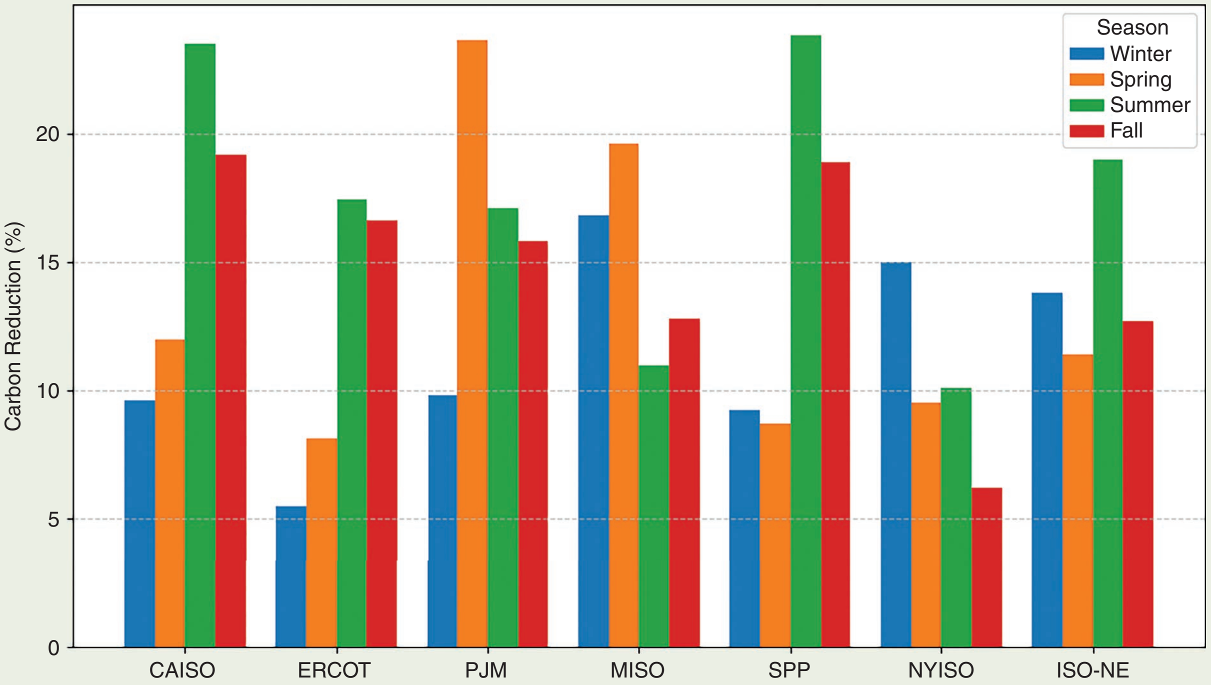


figure 5. Seasonal carbon emissions reduction across ISOs. The results indicate that carbon benefits vary more by season than by ISO, with summer providing the largest reductions across multiple regions.

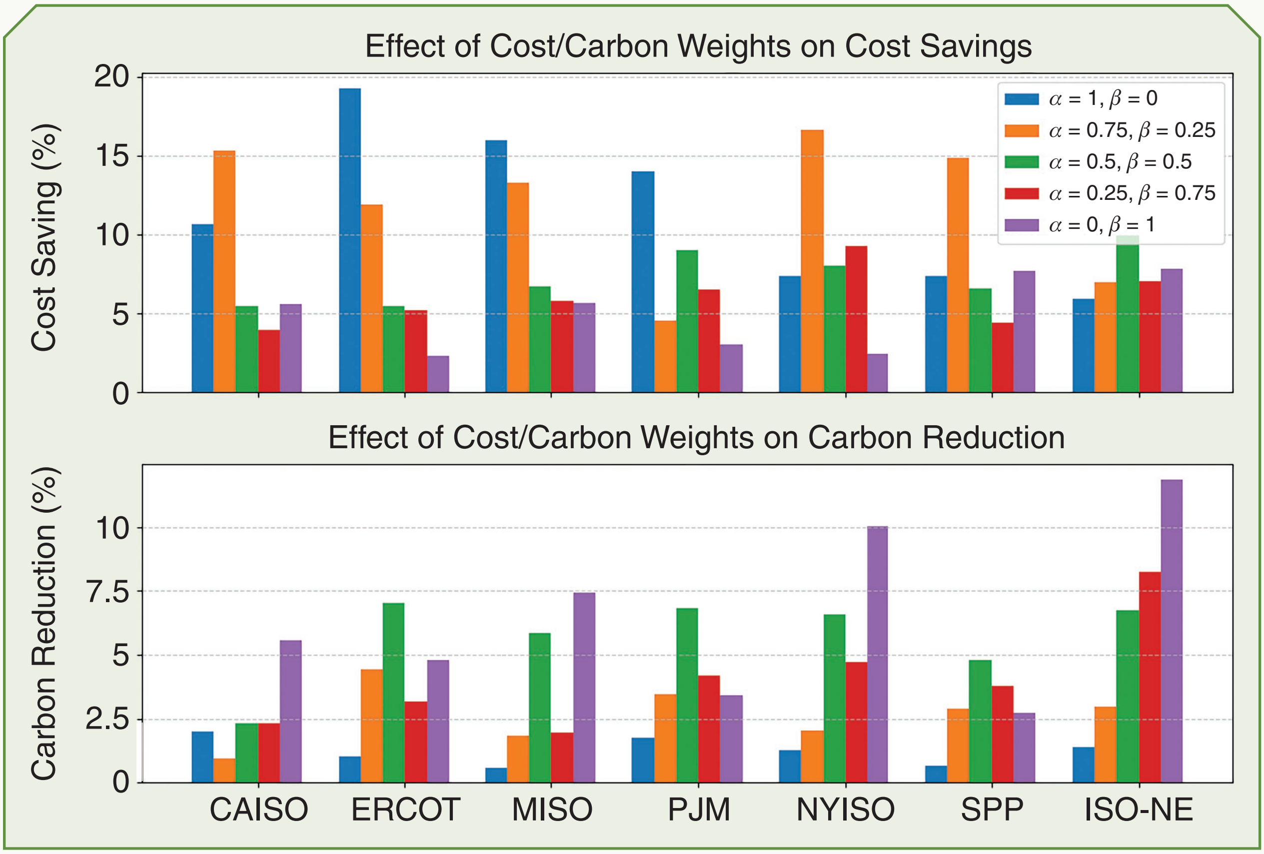


figure 6. The effect of cost–carbon weightings on workload migration benefits across U.S. ISOs. Higher cost weights maximize savings, while higher carbon weights achieve greater emissions reductions.

coal utilization. ERCOT, MISO, and ISO-NE maintain moderate but consistent reductions across seasons, while NYISO displays its weakest carbon benefit in fall, coinciding with lower renewable output and higher fossil dependence. Overall, carbon-driven migration opportunities are maximized during summer and, to a lesser extent, spring, when renewable generation is highest and interregional carbon intensity differentials are most pronounced.

Impact of Objective Weighting on Migration Benefits

Figure 6 shows how shifting the balance between cost and carbon objectives changes migration outcomes, with grid strain enforced as a hard constraint. Cost-heavy configurations deliver the largest cost savings but yield smaller carbon reductions. On the other hand, carbon-focused weights reverse this trend, achieving the highest emissions cuts in cleaner grids like CAISO and NYISO but with lower cost gains. Balanced weights produce moderate, well-rounded benefits, showing that small changes in priorities can significantly improve sustainability outcomes without sacrificing too much on cost. Although expected, these results highlight the importance of ability to have explicit control over optimization priorities. Interestingly, cost savings do not increase monotonically with higher cost weights (α) across all ISOs. This behavior reflects the nonconvex tradeoff surface between cost and carbon objectives under real market dynamics. In regions where electricity price and carbon intensity are weakly correlated (e.g., ERCOT and MISO), intermediate weighting schemes can lead to suboptimal migration decisions that minimize neither cost nor carbon effectively.

Comparison Against Existing Strategies

To evaluate the effectiveness of our workload migration framework, we compare against three baseline strategies that represent existing approaches in the literature, along with our proposed model. The baselines are designed to isolate the effect of optimizing for a single objective versus incorporating multiple objectives under operational constraints.

- *Cost-only*: Selects the migration destination that minimizes electricity cost without considering carbon intensity or grid strain. This represents cost-driven optimization policies commonly employed in cloud scheduling frameworks.

- *Carbon only*: Selects the migration destination that minimizes carbon intensity without considering electricity cost or grid strain. This reflects the design of carbon-aware schedulers that disregard electricity cost tradeoffs.
- *CarbonClipper*: Implements the migration strategy proposed by Lechowicz et al. (2024), which prioritizes carbon-aware migration while incorporating limited price awareness. CarbonClipper focuses on selecting low-carbon regions based on real-time signals with performance guarantees.
- *Our method*: Optimizes for a weighted combination of electricity cost and carbon intensity while enforcing strict constraints on grid strain and migration performance.

For each baseline, we use identical hourly electricity price, carbon intensity, and grid capacity datasets from major U.S. ISOs to ensure a fair comparison. All approaches are evaluated across the same workload traces and migration penalty assumptions. The cost-only and carbon-only baselines provide insight into the tradeoffs between optimizing for a single metric and jointly optimizing multiple objectives. The CarbonClipper baseline serves as a state-of-the-art reference for carbon-aware migration that has been demonstrated on real cloud workloads.

Figure 7 summarizes the results in terms of average carbon reduction, cost reduction, and percentage of intervals with grid strain violations across the evaluation period. While cost-only achieves the highest cost reduction, it also causes the highest rate of grid strain violations. On the other hand, carbon-only yields the largest carbon reduction but suffers from high operational costs and nonnegligible strain violations. CarbonClipper balances these two objectives better, but without explicit strain constraints, violations remain significant. Our model achieves competitive cost and carbon reductions while maintaining the lowest grid strain violation rate, demonstrating the importance of incorporating grid capacity constraints into the optimization process. Grid strain violations occur due to changes in demand and capacity. Since the optimization relies on day-ahead predictions of demand and capacity, unexpected spikes in real-time load can cause the actual strain to exceed modeled thresholds. Overall, our method achieves up to 15% cost reductions and 13% carbon reductions, while keeping the grid strain violations below 2.4%, showcasing the benefits of multiobjective optimization with grid strain constraints.

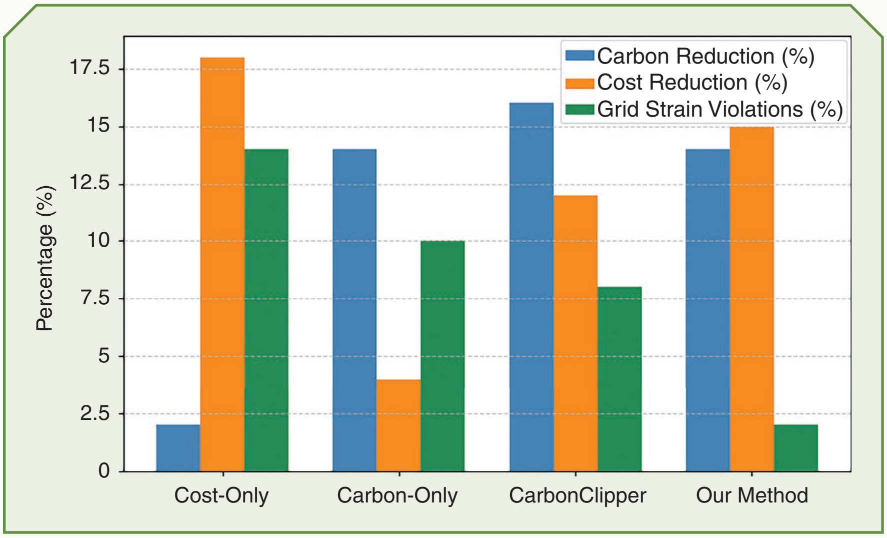


figure 7. A comparison of average carbon reduction, cost reduction, and grid strain violations across baseline strategies and our proposed model. Our approach achieves balanced cost and carbon savings while maintaining the lowest grid strain violation rate.

Key Takeaways

Our experiments show that moving computing workloads between regions can bring measurable benefits, but the tradeoffs are complex and require careful planning. When conditions are ideal, for example, when electricity prices are low and clean energy is available in the destination region, workload migration can reduce electricity costs by up to 59.9% and lower carbon emissions by as much as 62.8%. These savings demonstrate the strong potential of using location-based decisions to improve both economic and environmental outcomes. However, workload migrations without considering grid conditions can result in unexpected outcomes. Some regions may already be under stress, especially during peak demand periods, and adding more computational load to those areas can increase the risk of grid instability. Thus, it is essential to

> Our experiments show that moving computing workloads between regions can bring measurable benefits, but the tradeoffs are complex and require careful planning.

account for grid strain when planning where to send workloads.

In addition, we find that the effectiveness of migration strategies changes depending on the season and the specific regions involved. For instance, a migration path that works well in spring might not be as beneficial in the summer due to changes in electricity demand and the availability of renewable energy. This means that simple, one-size-fits-all migration rules are often not enough. Instead, operators should rely on flexible and dynamic strategies that adapt to real-time data about prices, emissions, and grid stress to make the best possible decisions. These findings motivate the need for fine-grained, real-time migration frameworks that consider carbon and cost as joint objectives, while ensuring grid reliability is not affected by aggressive migration decisions.

Limitations and Discussion

Our study presents a workload migration framework that minimizes electricity cost and carbon emissions while enforcing grid strain and workload performance as hard constraints. Using real-world datasets and seasonal simulations, we demonstrate that data center operators can achieve substantial economic and environmental gains by strategically relocating workloads while ensuring that the destination regions remain within safe grid operating limits.

One of the most important takeaways from our results is the interplay between the two optimization objectives (cost and carbon) and the operational constraints (grid strain and performance). Migrations that minimize electricity cost do not always coincide with those that reduce carbon emissions, particularly when lower-cost electricity is generated from fossil fuels. In addition, the grid strain constraint can restrict certain migrations even when they offer strong cost or carbon benefits, especially if the destination region is already operating near capacity. These results emphasize that the "optimal" migration policy depends not only on operator priorities but also on real-time operational constraints.

Our framework currently incorporates latency penalties as a soft factor in the objective function, but grid strain is modeled as a hard constraint. The results show that even modest latency penalties (e.g., 2.5%–5%) can reduce potential cost and carbon benefits, underscoring the importance of workload classification. Latency-tolerant batch workloads are prime candidates for migration, while real-time or interactive workloads may require additional strategies such as caching, prefetching, or partial migration to satisfy performance requirements.

Furthermore, our grid strain model currently captures only capacity constraints, without modeling transmission network congestion. As a result, the framework may overestimate feasible migration potential in regions with significant transmission bottlenecks. The current latency model assumes that migration delay scales linearly with workload size and inversely with bandwidth. Although this provides a reasonable first-order estimate, future work should incorporate the effects of traffic and routing congestion, which can affect end-to-end migration times. Our proposed method utilizes linear programming to solve the optimization problem. While linear programming ensures scalability and faster solutions, it simplifies certain aspects of the problem, such as nonlinear grid behaviors and price dynamics. Future work could explore mixed-integer or nonlinear formulations to capture network constraints or congestion effects more accurately. Future extensions of this work could include the following:

- integrating real-time forecasts and uncertainty models to enhance decision making under variable market and grid conditions
- developing seasonal migration profiles for specific workload types to exploit recurring cost and carbon advantages
- extending the model to intercontinental migrations to leverage time zone and seasonal generation differences (e.g., the United States versus Australia).

Conclusion

This article presents a framework for optimizing workload migration across regional data centers with the dual goals of reducing electricity costs and carbon emissions, while strictly enforcing operational constraints on grid strain and workload performance. By integrating real-world electricity price, carbon intensity, and grid capacity utilization data into a linear programming model, we enable informed migration decisions that deliver economic and environmental benefits without compromising grid reliability.

Our results demonstrate that strategic migration can achieve up to 15% electricity cost savings and 13% carbon reduction, provided that no destination region exceeds safe operating thresholds. Enforcing grid strain and performance constraints ensures that benefits are realized without shifting risk to other parts of the grid, making the approach both sustainable and operationally viable. Furthermore, seasonal and temporal variability in market and carbon signals highlight the value of dynamically adapting migration decisions over time. The modular nature of our framework allows it to accommodate different operator priorities, integrate additional constraints, and adapt to evolving grid conditions, making it a practical tool for sustainable, constraint-aware workload migration in geodistributed infrastructures.

Acknowledgment

This work was partially funded by the Research Council of Norway (L2D, award 344115), the Hariri Focused Research Program, and the NSF DESC program (award 2450111). Can Hankendi is the corresponding author.

For Further Reading

"Hyperscale data center capacity to triple by 2030, driven by generative AI," *Synergy*, Jan. 9, 2025. Accessed: May 12, 2025. [Online]. Available: https://www.srgresearch.com/articles/hyperscale-data-center-capacity-to-triple-by-2030-driven-by-generative-ai

T. Haslak, "Controllable loads in the smart grid: Stress relief for the power grid through rescheduling; Test implementation in sewage treatment facility," in *Proc. IEEE 11th Int. Symp. Appl. Comput. Intell. Inform. (SACI)*, 2016, pp. 381–384, doi: 10.1109/SACI.2016.

"Fundamentals of power grid reliability and clean electricity," Nat. Renewable Energy Lab., Tech. Rep. NREL/TP-6A40-85880, 2024. Accessed: May 17, 2025. [Online]. Available: https://www.nrel.gov/docs/fy24osti/85880.pdf

"OASIS market data." California Independent System Operator, 2025. Accessed: May 17, 2025. [Online]. Available: https://oasis.caiso.com/

"Market prices," Electric Rel. Council of Texas, Austin, TX, USA, 2025. Accessed: May 17, 2025. [Online]. Available: https://www.ercot. com/mktinfo/prices

"Electricity price ticker," PJM Interconnection, Valley Forge, PA, USA, 2025. Accessed: May 17, 2025. [Online]. Available: https://www.pjm.com/about-pjm/learning-center/price-ticker.aspx

"MISO live dashboard and price map," GridStatus.io, Chicago, IL, USA, 2025. Accessed: May 17, 2025. [Online]. Available: https://www.gridstatus.io/live/miso

"SPP live dashboard and price map," GridStatus.io, Chicago, IL, USA, 2025. Accessed: May 17, 2025. [Online]. Available: https://www.gridstatus. io/live/spp

"Energy market and operational data," New York Independent Syst. Operator, Rensselaer, NY, USA, 2025. Accessed: May 17, 2025. [Online]. Available: https://www.nyiso.com/energy-market-operational-data

"ISO express - Real-time maps and charts," ISO New England, Holyoke, MA, USA, 2025. Accessed: May 17, 2025. [Online]. Available: https://www.iso-ne.com/isoexpress

"Electricity maps: The world's most comprehensive electricity data platform." Electricity Maps, 2024. Accessed: May 17, 2025. [Online]. Available: https://www.electricitymaps.com

A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019. [Online]. Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

"Managing the seasonal variability of electricity demand and supply," Int. Energy Agency, Tech. Rep., 2023. [Online]. Available: https://www.iea.org/reports/ managing-the-seasonal-variability-of-electricity-demand-and-supply

A. Radovanović *et al.*, "Carbon-aware computing for datacenters," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1270–1280, Mar. 2023, doi: 10.1109/TPWRS.2022.3173250.

A. Lechowicz *et al.*, "CarbonClipper: Optimal algorithms for carbon-aware spatiotemporal workload management," 2024, arXiv:2408.07831.

Biographies

Can Hankendi is with Electrical and Computer Engineering Department, Boston University, Boston, MA 02215 USA.

Ayse K. Coskun is with Electrical and Computer Engineering Department, Boston University, Boston, MA 02215 USA.