Electric Power Systems Research 212 (2022) 108586

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Using geographic load shifting to reduce carbon emissions

Julia Lindberg *, Bernard C. Lesieutre, Line A. Roald

University of Wisconsin-Madison, United States

ARTICLE INFO

Keywords:
Load shifting
Carbon reduction

ABSTRACT

An increasing focus on the electricity use and carbon emissions associated with computing has lead to pledges by major cloud computing companies to lower their carbon footprint. Data centers have the unique ability to shift computing load between different geographical locations, giving rise to flexibility that can be employed to reduce carbon emissions. In this paper, we present a model where data centers shift load independently of the ISOs. We first consider the impact of load shifting guided by locational marginal carbon emissions, λ_{CO_2}, a sensitivity metric that measures the impact of incremental load shifts. Relative to previous models for data center load shifting, the presented model improves accuracy and includes more realistic assumptions regarding the operation of data centers and electricity markets. Further, we introduce a benchmark model where data centers have access to the full information about the power system and can identify optimal shifts for the current time period. We demonstrate the efficacy of our model on the IEEE RTS-GMLC system using 5 min load and generation data for an entire year. Our results show that the proposed improvements for the shifting model based on λ_{CO_2} are highly effective, leading to results that outperform the benchmark model.

1. Introduction

The recent technology revolution has infiltrated most aspects of modern day life, and has lead to an increase in demand for computing resources. Between 2010 and 2018 there was an estimated 550% increase globally in the number of data center workloads and computing instances [1]. This increased computing demand has commanded a shift from smaller data centers to large-scale, highly optimized and efficient facilities. These facilities are referred to as *hyperscale* data centers and are operated by large technology companies such as Amazon, Facebook, Google, Microsoft and Alibaba. These companies operate vast networks of these data centers which are dispersed geographically throughout the world [2,3]. Data centers currently consume around 1%–2% of electricity globally [1]. This low percentage is largely due to the increased efficiency of hyperscale data centers, which have processed the increased computing demand while only increasing electricity use by an estimated 6% [1]. However, as the scope for further efficiency gains is largely exhausted, it is projected that electricity use for computing will increase rapidly in the future.

Hyperscale data centers are large loads on electric power networks with unique characteristics. For example, data centers can choose to defer when computing tasks are processed or process them at different locations. These properties equip data center operators with the unique ability to participate in both geographic and temporal load shifting. Motivated by recent pledges made by technology companies to reduce their carbon emissions [4–6], we are interested in understanding how these hyperscale data centers can effectively interact with electricity markets in a way that minimizes the carbon emissions caused by their computing loads. Companies such as Google already employ carbon aware computing systems [7] and recent start-ups consider computing that adapts to the operation of the electric grid [8], an idea that plays a critical role in the vision of zero carbon cloud computing [9,10].

When shifting load, data centers interact with electricity markets. The metrics used to determine the best shifting decisions has a significant impact on the market outcomes [11]. In this paper, we focus on areas with central-dispatch electricity markets cleared using a DC OPF, representative of areas operated by ISOs in the United States. Previous research has examined the impact of integrating data centers and demand response [12–18] or have considered geographical load shifting to reduce electricity costs [19–21], including load shifting between different electricity markets [22]. Other work has modeled data center flexibility through the use of virtual links in time and space [23]. Much of this body of work considers the shifting of computing load to reduce the carbon emissions of data centers [21,24,25] and increase absorption of renewable energy [26,27]. There are, however, several major challenges to effective collaboration between the ISO and data centers, such as barriers to the exchange of possibly sensitive information. More fundamentally, ISOs and data centers have different objectives, namely, ISOs focus on minimizing electricity cost while data

* Corresponding author.
E-mail address: jrlindberg@wisc.edu (J. Lindberg).

https://doi.org/10.1016/j.epsr.2022.108586
Received 3 October 2021; Received in revised form 17 April 2022; Accepted 2 July 2022
Available online 23 July 2022

centers would like to minimize carbon emissions. While it is a common assumption that those two objectives are well aligned, our previous research [11] found that if data centers are specifically targeting the objective of reducing carbon emissions (rather than reducing cost) they can achieve better results by doing the load shifting outside of the electricity market rather than collaborating with the ISO. Since ISOs are unlikely to change the market to directly minimize carbon emissions in the near future, we focus on a method to identify geographical load shifts that reduce carbon emissions within the existing electricity market.

To effectively shift load geographically, we require information about the locational variation in the carbon footprint of electricity, similar to the way locational marginal prices describe the locational variations in the cost of electricity. While locational marginal electricity prices are made publicly available, similar locational data on the carbon emissions associated with electricity usage is generally not. As a result, it is less straight forward to develop metrics that provide good guidance for shifting load to reduce carbon emissions. Previous work has assumed that prices are directly tied to the fraction of non-renewable energy [28], or considered average carbon emissions for electricity in a region and/or renewable energy curtailment [27,29,30]. A benefit to these metrics is that several companies provide the necessary information needed to compute them [31–33]. A downside is that they fail to reflect the impact of a marginal change in load. Specifically, there is no consideration for which generators in the system will be providing the additional generation, which depends on factors such as generation and transmission capacity limits. To more effectively guide load shifting, we need a measure of marginal emissions [11]. The use of marginal emissions to assess energy efficiency or the potential impact of renewable energy has been considered in the academic literature [34,35] and by industry [36]. They have also been proposed to guide renewable energy investments [34,37,38], or to assess the impact of marginal emissions rates at cogeneration facilities [39]. Other work has considered calculating marginal carbon emissions via LMPs [40] and used marginal carbon emissions to explore an alternative load distribution management paradigm [41]. More recently, PJM has started to make marginal carbon emissions publicly available [42–44].

Locational marginal carbon emissions was recently proposed to guide data center geographical load shifting [45]. The superiority of this metric over others such as average carbon emissions, curtailment or LMPs for data center geographic load shifting was demonstrated in [11]. While the previous load shifting models in [11,45] presented promising results, it included several unrealistic assumptions. First, [11, 45] assumed that all computing loads are reassigned in each time step and that the electricity market is being resolved twice to account for load shifting. Here, we address this issue by implementing a cumulative load shifting model and calculating the load shifts for the current time-step based on (the already known) locational marginal emissions from the previous time-step. Further, due to the fact that the locational marginal emissions as derived in [45] represent a local sensitivity factor, the metric may not be accurate for large load shifts and can fail to find the optimal shifts. Therefore, we present a benchmark model based on bilevel optimization to assess the level of inaccuracy and effectiveness of this metric. We also propose to use regularization (which can be interpreted as imposing a cost on load shifting) to increase the accuracy of the model. In summary, the three main contributions of this paper are the following:

(1) Realistic model for data center load shifting: We address the unrealistic assumptions in our previous model by implementing load shifting in a cumulative fashion and incorporating a regularization parameter to increase the accuracy of shifting.

(2) Benchmark model for optimal data center shifting: We develop a new benchmark model that computes the optimal data center load shift, given access to all system information. This bilevel optimization method determines the optimal load shift to reduce carbon emissions, subject to the operation of the current electricity market.

(3) Computational Analysis: We compare both models using the RTS-GMLC system with one year of 5 min load and generation data, and observe that the shifting model based on locational marginal carbon emissions performs quite well.

The remainder of the paper is organized as follows. In Sections 2 and 3 we describe how to model data center load flexibility and we review the existing data center driven shifting model. Section 4 presents the proposed improvements to this model. In Section 5, we present a new benchmark model for optimal load shifting. Section 6 demonstrates the efficacy of our model in a case study, and Section 7 concludes.

2. Model of data center load shifting

The ability of data centers to shift load is highly dependent on the characteristics of the workload (i.e., the stream of computational tasks that the data center has to process). Important considerations include access to the necessary data (e.g., the input data needed to perform a certain task may only be available certain locations), hardware efficiency (e.g., certain data centers may have access to tailored hardware that allow them to perform certain kinds of computational tasks more efficiently) and requirements on latency (i.e., how quickly a computing task can be completed). Determining the inherent flexibility of a workload and an associated network of data centers is a topic of ongoing research, and requires access to proprietary data. In this paper, we do not attempt to actually schedule data center workloads, but rather develop an abstracted model assuming that we have access to some high-level information about the aggregate ability of a data center to shift load.

Furthermore, in the following, we will assume that decisions on data center load shifts happens on a 5-min time scale. We note that this does not imply that individual loads are only allowed to run for 5 min and then risk to be shifted to another data center. We assume that a sufficient number of computing tasks are completed or arrive within 5 min such that the overall data center shifting flexibility is primarily determined by the scheduling of newly incoming tasks.

We next discuss our model of the data center load shifting capabilities and load shifting cost which aims to capture the key characteristics of data center flexibility.

2.0.1. Load shifting capabilities

Following the model in [45], we let C denote the set of data center loads that can geographically shift load. We denote the change in load at data center i by $\Delta P_{d,i}$, and the shift in load from data center i to j by s_{ij}. Then, the load shifting capabilities can be represented using the following constraints

$$\Delta P_{d,i} = \textstyle\sum_{j\in C} s_{ji} - \sum_{k\in C} s_{ik} \qquad \forall i \in C \tag{1a}$$

$$\sum_{i\in C} \Delta P_{d,i} = 0 \tag{1b}$$

$$-\epsilon_i \cdot \mathrm{Cap}_i \le \Delta P_{d,i} \le \epsilon_i \cdot \mathrm{Cap}_i \qquad \forall i \in C \tag{1c}$$

$$0 \le \Delta P_{d,i} + P_{d,i} \le \mathrm{Cap}_i \qquad \forall i \in C \tag{1d}$$

$$0 \le s_{ij} \le M_{ij} \qquad \forall ij \in C\times C. \tag{1e}$$

Here, the constraint (1a) enforces that the change in load at a given data center is equal to the total load shifted in minus the total load shifted out, while (1b) says the sum of all load shifts across all data centers must be zero. We note that (1a), (1b) implicitly assume that the efficiency of each data center is the same, i.e. the electric power used to process a given task is the same even if the task is shifted from data center i to j. However, this may not be the case as some data centers may be better equipped (i.e., have specialized hardware or easier access to data needed for a computation) and thus be able to perform a certain computing task with a smaller amount of power. Such differences in efficiency can be incorporated as losses in transmitting the computing load from more to less efficient data centers in (1a) and considering a weighted sum of $\Delta P_{d,i}$ where the weight gives a measure of the efficiency of the data center in (1b). However, for simplicity (and a lack of data on this behavior) we assume that all data centers are equally efficient. Constraint (1c) limits the amount of computing load each data center can shift. While this limit is workload dependent and should be defined separately for each data center, we express it as a fraction ϵ of the data center capacity Cap for simplicity. Constraint (1d) enforces the total computing load at each data center i to remain below the total capacity Cap_i and (1e) limits how much load data center i can send to data center j to remain below M_{ij}.

2.0.2. Load shifting cost

In addition to the ability of data centers to shift load described above, we can include a cost on data center load shifting. We consider this cost to be expressed as a quadratic penalty on the total load shifts, i.e.,

$$\text{Cost of total shift} = \gamma \|\Delta P_{d,i}\|_2^2 \tag{2}$$

This cost can also be interpreted as a regularization term, and the effect of incorporating this cost will be discussed below.

3. Review of existing load shifting model

Previous work has considered ISO independent load shifting to reduce carbon emissions via a value called the *locational marginal carbon emissions*, λ_{CO_2} [45]. This value is calculated as a linear sensitivity around the optimal solution to the DC OPF. We give a brief outline of this load shifting model here, but suggest [45] as a more detailed reference.

Step 1: This model begins by assuming the ISO solves a DC OPF [46, 47]. The DC OPF is a linear optimization problem that seeks to minimize generation costs subject to network and demand constraints with decision variables $x = [\theta \ P_g]$ where P_g are the generation variables, θ are the voltage angles at each node. Let $\mathcal{N}$ be the set of all nodes, $\mathcal{L}$ the set of lines and $\mathcal{G}$ the set of generators. The DC OPF is defined as

$$\min_{\theta, P_g} \quad c^T P_g \tag{3a}$$

$$\text{s.t.} \quad \sum_{\ell \in \mathcal{G}_i} P_{g,\ell} - \sum_{\ell \in D_i} P_{d,\ell} = \sum_{j:(i,j)\in\mathcal{L}} \beta_{ij}(\theta_i - \theta_j), \qquad \forall i \in \mathcal{N} \tag{3b}$$

$$-P_{ij}^{lim} \le -\beta_{ij}(\theta_i - \theta_j) \le P_{ij}^{lim}, \qquad \forall (i,j) \in \mathcal{L} \tag{3c}$$

$$P_{g,i}^{min} \le P_{g,i} \le P_{g,i}^{max}, \qquad \forall i \in \mathcal{G} \tag{3d}$$

$$\theta_{ref} = 0. \tag{3e}$$

The cost function (3a) minimizes the cost of generation, where c_i is the cost of generation at generator i. Constraints (3b)–(3d) constrain the nodal power balance, transmission line and generator capacity constraints respectively. Finally, (3e) sets the voltage angle at the reference node to zero.

Step 2: Independent of any ISO collaboration, data center operators shift their load to minimize carbon emissions. To guide this effort, a metric known as the *locational marginal carbon emissions* was proposed in [45]. In [11] the authors demonstrated the superiority of this metric to other more commonly studied metrics such as the average carbon emissions or excess low carbon power.

To calculate the locational marginal carbon emissions, we first consider a basic optimal solution $x^* = [\theta^*, P_g^*] \in \mathbb{R}^n$ to the DC OPF (3). From sensitivity analysis in linear optimization theory [48], a basic optimal solution can be written as $Ax^* = b$, where $A \in \mathbb{R}^{n\times n}$ is a full rank matrix consisting of all the active constraints of (3) at the optimal solution x^*. In the case of the DC OPF, the rows of A consist of the equality constraints (3b) and (3e) as well as a subset of the inequality constraints (3c), (3d) that are satisfied at equality for x^*.

A small change in load can be represented as a small change in the right hand side b, given by $\Delta b = \begin{bmatrix}\Delta P_d & 0\end{bmatrix}^T$. Assuming that the change is small enough as not to alter the set of active constraints, we compute the optimal change in generation as $A \cdot \Delta x = \Delta b$ where $\Delta x = [\Delta\theta \ \Delta P_g]$. This gives rise to the linear relationship

$$\begin{bmatrix}\Delta\theta \\ \Delta P_g\end{bmatrix} = A^{-1} \cdot \begin{bmatrix}\Delta P_d \\ 0\end{bmatrix} \tag{4}$$

If we denote the matrix consisting of the last $|\mathcal{G}|$ rows and first $|\mathcal{N}|$ columns of A^{-1} by B, this gives the linear relationship between load and generation changes, $\Delta P_g = B \cdot \Delta P_d$.

Consider the cost vector $g \in \mathbb{R}^{|\mathcal{G}|}$ that measures the carbon emissions of each generator per MW. Specifically, the ith component of g, is the carbon intensity of generator i. Multiplying each side of $\Delta P_g = B \cdot \Delta P_d$ on the left by g gives us the change in carbon emissions

$$\Delta \text{CO}_2 = g \cdot \Delta P_g = g \cdot B \cdot \Delta P_d = \lambda_{\text{CO}_2} \Delta P_d. \tag{5}$$

where $\lambda_{CO_2} = g \cdot B$. Intuitively, the kth component of λ_{CO_2} measures how an increase of 1 MW of load at node k will affect the total carbon emissions of the system.

We let C denote the set of data center loads that can geographically shift load and consider optimization variables that denote the change in load at data center i, $\Delta P_{d,i}$, and the shift in load from data center i to j, s_{ij}. The geographic load shifting optimization problem is given by:

$$\begin{aligned} \min_{\Delta P_d, s} \quad & \sum_{i\in C} \lambda_{\text{CO}_2,i} \Delta P_{d,i} \\ \text{s.t.} \quad & \text{Data center shifting constraints (1)} \end{aligned} \tag{6a}$$

The objective value (6a) minimizes the change in carbon emissions as a function of the change in load, and the data center load shifting capabilities are expressed using the constraints (1). Note that in this original model, we do not include any cost penalty on load shifting.

Step 3: Finally, the ISO resolves the DC OPF (3) with new load profile, $P'_{d,i} = P_{d,i} + \Delta P^*_{d,i}$, where $\Delta P^*_{d,i}$ is the optimal solution to (6) for all $i \in \mathcal{N}$.

4. Realistic data center load shifting model

The above model has several drawbacks. First, it is unrealistic to assume that the ISO resolves the market clearing twice, once before and once after the shifting has happened. Second, since the model is linear, we tend to see large load shifts even with small differences in λ_{CO_2} between data center locations. Since λ_{CO_2} is a local sensitivity factor that is only accurate near the previous optimal solution, these large shifts lead to inaccurate results that sometimes increase carbon emissions. To address these issues, we introduce two improvements to the model: cumulative load shifting and regularization.

4.1. Cumulative load shifts

We refine the model defined in [45] by considering *cumulative load shifts*. Instead of resolving the DC OPF in Step 3 of the above model, the load shift is applied to the market clearing in the next time step. Specifically, the algorithm runs as follows:

Step 1: At time t, the ISO solves the DC OPF (3) with data center load set to P_d^t.

Step 2: Given information about λ_{CO_2} as described above, the data center operator computes a load shift ΔP_d^t according to (6). Then, the data center load for time $t+1$ is set to $P_d^{t+1} = P_d^t + \Delta P_d^t$, and the algorithm proceeds to Step 1 of the next time step.

While the cumulative load shifting model more accurately reflects the current market set up, it introduces an additional inaccuracy in our model. The locational marginal carbon emission value λ_{CO_2} at each data center is derived as a linearization from the operating point at time t, but the internal data center shifting optimization will only affect the market clearing at time $t+1$. The expectation is that since operating conditions remain similar between time steps, shifting with respect to λ_{CO_2} will still lead to a decrease in total system carbon emissions. We note that this may not be the case. In other words, marginal generators can change significantly from one time step to the next, but in general this is not the case.

We also note that cumulative load shifting can increase accuracy relative to the existing model, particularly the load shift allowed in each time step is small (i.e., only a small fraction ϵ can be shifted). In this case, changes in the data center load build up slowly over time. This is in contrast to our previous model, where the data center load was reset to the original value P_d in each time step.

4.2. Regularizing load shifts

To discourage large load shifts which can cause oscillations and increased emissions, we propose to use a regularization term (i.e., a quadratic penalty) that discourages large shifts. Specifically, this model replaces the objective value (6a) with

$$\sum_{i \in C} \lambda_{CO_2,i} \Delta P_{d,i} + \gamma \|\Delta P_{d,i}\|_2^2$$

where $\gamma \in \mathbb{R}$ is a regularization parameter. The goal in using this regularization term is to discourage large shifts that lead to an increase in carbon emissions as well as increase the accuracy of the data center driven shifting model. However, as discussed above, the regularization term can also be interpreted as a quadratic cost on load shifting. The use of a quadratic penalty on the shifts ensures that while small shifts are cheap and frequent, we only shift a large amount of load when there will be a large reduction in carbon emissions.

Throughout the rest of this paper we refer to the model outlined in this section as (λ_{CO_2}-shift).

5. Benchmark model for optimal shifting

Our next contribution is to introduce a new model to benchmark the data center driven shifting model. Since the shifts provided by λ_{CO_2} are calculated by a linear sensitivity, they can be inaccurate, even giving shifting profiles that increase carbon emissions.

The problem of identifying the optimal load shift data center operators should employ to minimize carbon emissions can be modeled as a bilevel linear program. The upper level problem identifies the optimal choice of load shift ΔP_d to minimize carbon where as the lower level problem ensures all network constraints are met. The optimal load shift is a solution to the bilevel optimization problem:

$$\begin{aligned} \min_{\Delta P_d, s, P_g^*} \quad & g^T P_g^* \\ \text{s.t.} \quad & P_g^* = \arg\min \text{ (DC-shift)} \\ & (\Delta P_d, s) \in \mathcal{P} \end{aligned} \tag{Opt-shift}$$

Here, the last constraint represents the set of feasible load shifts from the data center perspective, i.e., $\mathcal{P}$ is the polytope of permissible load shifts defined by the constraints in (1). The first constraint states that the generation values P_g^* is the solution to the lower level optimization problem (DC-shift). This problem is a version of the standard DC OPF (3) where the nodal balance constraints include the change in demand. Formally we write it as

$$\min_{P_g, \theta} \; c^T P_g \quad \text{subject to}$$

$$\text{Constraints (3c),(3d),(3e)} \tag{DC-shift}$$

$$\sum_{\ell \in \mathcal{G}_i} P_{g,\ell} - \sum_{\ell \in \mathcal{D}_i} (P_{d,\ell} + \Delta P_{d,\ell}) = \sum_{j:(i,j) \in \mathcal{L}} -\beta_{ij}(\theta_i - \theta_j), \qquad \forall i \in \mathcal{N}$$

As in (λ_{CO_2}-shift), we also consider cumulative load shifting in (Opt-shift). Specifically, at each data center $\ell \in C$, at time step t we assume the load $P_{d,\ell}$ in (DC-shift) reflects the sum of new load $P_{d,\ell}$ from time t and the load shift $\Delta P_{d,\ell}$ from time $t-1$.

6. Computational results

We next analyze the accuracy and efficacy in carbon reduction of (λ_{CO_2}-shift) versus (Opt-shift). We first discuss our test case and the performance evaluation criteria before presenting numerical results.

6.1. Test case

We perform an extensive year long analysis of carbon reduction methods mentioned above using the RTS-GMLC system [49]. This system has 73 buses, 158 generators and 120 lines. Since the original system does not designate data center loads, we assign data centers at buses 103, 107, 204 and 322. These buses were chosen as they had the smallest and largest λ_{CO_2} values for the single time step data given in [49]. We assume that cumulatively, the four data centers consume a fixed power of 1000 MW at each time step throughout the year, although the distribution of that power among the four data centers varies. We assume that the data center load is flat since due to the costs of capital investment and resource utilization, data centers typically run close to capacity. In addition, changes in load is slow, much slower than time scale of shifting and therefore is not relevant to our results. We note that our model could be run with a time-varying profile without any problems, we just use 1000MW for simplicity. We assume at time step 0, each data center starts with 250 MW of load. For all other loads and renewable generation, we use the real time, i.e. 5 minute, load and generation data provided with [49]. Over the course of the year, this system serves 526, 220, 000 MW of load, and 105, 408, 000 MW or roughly 20.03% of it is data center load.

Adding these large data center loads to the network greatly increases the total system load and results in time steps where the original DC OPF is infeasible. To remedy this we change the generation limits by setting $P_g^{\min} = 0$ for all $g \in \mathcal{G}$, and increase the maximum generation limits by 50%. At each time step we allow each data center to shift up to 20% of its total capacity, i.e. 50 MW, which implies that $\epsilon = 0.2$. We further enforce that data center capacities remain between 0 and 300 MW. We note that this shifting flexibility is large compared to existing capabilities of data centers [7], but it may be representative of future shifting capabilities and it demonstrates the potential effects of large-scale load shifting.

6.2. Performance evaluation

When evaluating performance of our methods, our most important metrics are reduction in carbon emissions and total cost of system operation across the year. When evaluating these metrics for (λ_{CO_2}-shift), we consider two different metrics:

Predicted cost and emissions: By evaluating $\Delta P_g = B \cdot \Delta P_d$, we obtain the predicted change in generation as a result of the load shift ΔP_d. Using this predicted generation change, we can derive a value for the predicted change in cost and carbon emissions to the system as a result of the load shift.

Actual cost and emissions: The actual carbon cost and carbon emissions are obtained from the DC OPF solved by the ISO after the load shift.

We note that these two metrics are different because λ_{CO_2} is only a local sensitivity factor that may be inaccurate for large shifts, and the load shift in one time step is calculated based on the λ_{CO_2} in the previous time step, which may no longer be accurate. In contrast, the carbon emissions and generation costs obtained with the benchmark model directly reflect the actual cost and emissions.

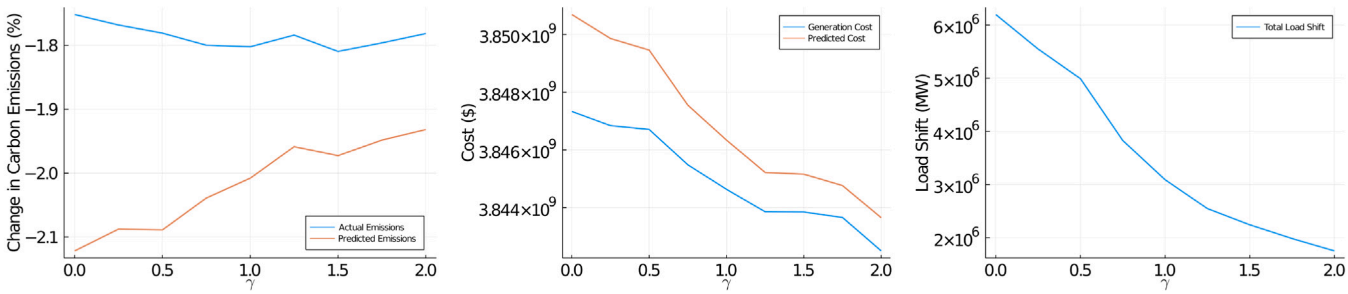

Fig. 1. Change in carbon emissions (left), generation cost (middle) and load shift (right) as γ varies.

Table 1
Summary of results from all models.

	DC OPF	(Opt-shift)	λ_{CO_2}-shift: $\gamma = 0$	λ_{CO_2}-shift: $\gamma = 1.5$
Generation cost	3, 802, 706, 000	4, 981, 076, 000 (30.99%)	3, 847, 332, 000 (1.17%)	3, 843, 847, 000 (1.08%)
CO_2 emissions	164, 402, 000	110, 444, 000 (−32.82%)	161, 522, 000 (−1.75%)	161, 427, 000 (−1.81%)
Total shifts	0	1, 048, 000	6, 199, 000	2, 245, 000

6.3. The effect of regularization

We first investigate the effect of the regularization parameter γ. The effect of various regularization parameters on generation cost, total system carbon emissions and total load shift is shown in Fig. 1, where the orange and blue lines represent the predicted and actual values, respectively. Fig. 1shows that the minimum total system carbon emissions occurs when the regularization parameter $\gamma = 1.5$ is used. In addition, we see that as the regularization parameter γ increases, the difference between the predicted carbon emissions and generation cost and the actual carbon emissions and generation costs decreases. This indicates that including a regularization term helps not only in the efficacy of the data center driven shifting model, but also in the accuracy.

The reason regularization is considered is to discourage load shifting in cases where it is not predicted to make large differences. The rightmost plot in Fig. 1 shows how as the regularization parameter increases, the total load shifted throughout the year decreases. We see that when the regularization parameter is set at $\gamma = 1.5$, the total amount of load shifted is less than half of the amount of load shifted when $\gamma = 0$. Considering that the carbon emissions and generation cost when $\gamma = 1.5$ are lower than when $\gamma = 0$, this demonstrates that shifting less load, more strategically can lead to a larger reduction in carbon emissions and a smaller increase in generation costs.

6.4. Comparison with opt-shift and original DC OPF solution

We next compare the solutions for (λ_{CO_2}-shift) with regularization parameters $\gamma = 0$ and $\gamma = 1.5$ with the original DC OPF solution and the solution obtained using our benchmark model (Opt-shift). These results are given in Table 1. We see that when considering (λ_{CO_2}-shift) with no regularization, carbon emissions relative to the original DC OPF decreases by around 2.8 million tons or 1.75%. This reduction is achieved while shifting around 6.2 million MW of load. Conversely, once the regularization term $\gamma = 1.5$ is added, we achieve an even greater reduction in carbon emissions, namely 2, 975, 000 tons or 1.81% while only shifting around 2.25 million MW of load. In addition, when considering regularization, total system generation costs only increased by 1.08% while without regularization it increased by 1.17%.

In contrast to the above results, we see a dramatic carbon savings when using the benchmark (Opt-shift). In this case we save 53, 958, 000 tons of carbon, i.e. 32.82%. This occurs while only shifting a little over 1 million MW. This dramatic savings occurs at a major increase to generation costs. Namely, (Opt-shift) results in an increase in \$1, 178, 370, 000 to generation costs or 30.99% over the original DC OPF.

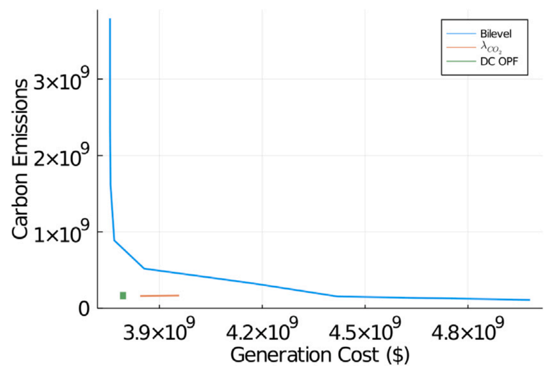

Fig. 2. Trade off between carbon emissions and generation cost.

This benchmark model suggests that dramatic reductions in carbon emissions are possible even with limited data center flexibility, but come at a large increase to generation costs.

6.5. Carbon emissions vs generation costs

As seen above, minimizing carbon emissions can lead to an increase in generation cost. To better understand the trade-off between carbon emissions and cost, we consider the benchmark model (Opt-shift) with objective function

$$(\alpha c^T + (1-\alpha) g^T) P_g^*$$

and (λ_{CO_2}-shift) with objective function

$$(\alpha \mathrm{LMP} + (1-\alpha)\lambda_{CO_2}) \Delta P_d + 1.5 \cdot \|\Delta P_d\|_2^2$$

in place of (6a) where $\alpha \in [0, 1]$ is a trade off parameter that allows us to weight the emphasis on minimizing carbon emissions versus generation costs and LMP is a vector of the locational marginal prices at each node. The trade off between minimizing carbon emissions and generation cost is shown graphically in Fig. 2.

When considering (λ_{CO_2}-shift), shown in orange, we see a small variation in the overall system generation cost and carbon emissions that remains close to the carbon emissions and generation cost of the original DC OPF. This is consistent with the results shown above, and is due to the fact that this model considers small shifts away from an operating point that minimizes generation costs. The benchmark model (Opt-shift) produces a much larger variation in operating points as we change the trade-off parameter α. As α increases, the model produces a

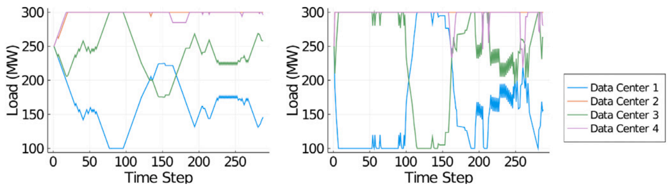

Fig. 3. Load at each data center during the first 24 hours using (λ_{CO_2}-shift) with $\gamma = 1.5$, $\epsilon = 0.01$ (left) and $\epsilon = 0.2$ (right).

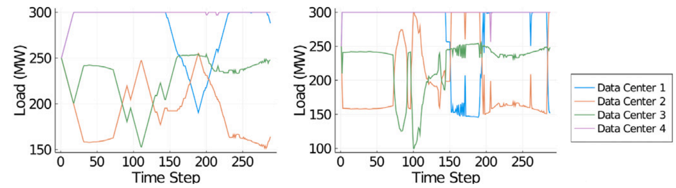

Fig. 4. Load at each data center during the first 24 hours using (Opt-shift) with $\epsilon = 0.01$ (left) and $\epsilon = 0.2$ (right).

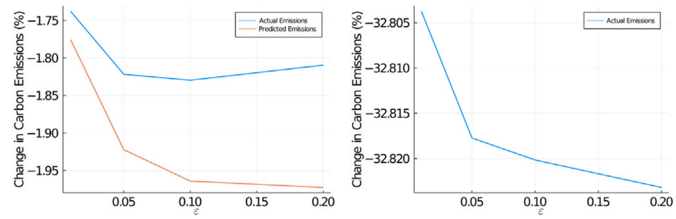

Fig. 5. Predicted and actual change in carbon emissions using (λ_{CO_2}-shift) (left) and the change in carbon emissions using (Opt-shift) (right) for varying epsilon values.

large increase in carbon emissions for only a moderate cost savings. In addition, we see that for (Opt-shift) to achieve lower carbon emissions than the DC OPF and the (λ_{CO_2}-shift), a major increase to generation cost is needed. This demonstrates that even with limited geographic load shifting flexibility, a large reduction in carbon emissions is possible but it comes at the price of significantly higher generation costs.

Fig. 2 also demonstrates an interesting phenomenon, namely the greedy nature of (Opt-shift). When only trying to minimize carbon emissions, (Opt-shift) is able to reduce total system carbon emissions by roughly 33% but this comes at a significant increase to total system generation cost. However, for the same generation cost, (Opt-shift) gives a solution with higher carbon emissions than the DC OPF or (λ_{CO_2}-shift). This demonstrates that the greedy nature of (Opt-shift) is not necessarily an optimal way to shift load over a long time span. Specifically, (Opt-shift) finds a load shift that gives the largest reduction in carbon emissions at that time step, with no consideration to how the load shift will affect the carbon emissions of the system at the next time step. Using forecasts of future load and generation information to aid in a long term load shifting strategy and developing a model that considers dependencies between time steps is left as future work.

6.6. Data center operating load

Finally, we consider the impact of each model on the data center operating load. We consider (λ_{CO_2}-shift) with regularization parameter $\gamma = 1.5$ and (Opt-shift), and two different limits on the amount of load that can be shifted in each time step, $\epsilon = 0.01$ and $\epsilon = 0.2$.

In Fig. 3 we see the operating conditions of each data center over the course of the first day when using (λ_{CO_2}-shift) when $\epsilon = 0.01$ (left) and $\epsilon = 0.2$ (right). In both cases we see similar overall trends in operating load. However, $\epsilon = 0.2$ leads to much quicker changes and also dramatic oscillations in the load at data centers 1 and 3 towards the end of the day. Similarly, in Fig. 4 we see the operating conditions of each data center over the course of the first day using (Opt-shift) when $\epsilon = 0.01$ (left) and $\epsilon = 0.2$ (right). Again, we see similar trends data center load for both values of ϵ, but for (λ_{CO_2}-shift), $\epsilon = 0.2$ leads to more oscillations in data center operating load.

Interestingly, there are some differences between the (λ_{CO_2}-shift) and (Opt-shift). In both cases we see an initial pull for data center 4 to operate at maximum capacity while the other data centers operate at lower capacities. This implies that the λ_{CO_2} value for data center 4 is accurately dictating that it is the most carbon neutral data center. In contrast, we see that when shifting with respect to (λ_{CO_2}-shift), data center 2 is also operating at maximum capacity. This is in contrast to shifting when using (Opt-shift). In this instance data center 2 initially drops to be the data center operating at the lowest load. This discrepancy highlights the inaccuracy when shifting with respect to λ_{CO_2}.

Finally, we investigate how using different ϵ values impacts the overall effect on carbon emissions. In Fig. 5 we see the change in total system carbon emissions as ϵ varies for both (λ_{CO_2}-shift) with $\gamma = 1.5$ as well as (Opt-shift). In both cases we see only a very mild decrease in total carbon emissions as we allow ϵ to increase. Further, for (λ_{CO_2}-shift), we see that as ϵ increases, the accuracy of the model decreases and once $\epsilon > 0.1$, the carbon emissions starts to increase. This indicates that allowing small shifts not only is more desirable from an operational stand point to avoid rapid changes and oscillations in data center loading, but it leads to similar carbon savings as allowing larger shifts.

7. Conclusion

In this paper we presented an improved model for data center load shifting to reduce carbon emissions. This model shifts load independently of ISO collaboration via a measure known as locational marginal carbon emissions. We built on existing work, but made several improvements to increase realism and accuracy of our model. We also proposed a new benchmark model which gives the best load shift at each time step, and compared the results. The main conclusion from the paper is that smaller load shifts, limited by regularization and shifting caps, are quite effective in reducing carbon emissions. Larger load shifts tend to decrease accuracy of the model and produce less carbon savings. Further, while our benchmark model is able to achieve large carbon reductions, it also significantly increases cost.

This paper demonstrates many natural directions for future work. First, this work shows that shifting load in a greedy way, i.e. shifting to reduce the maximal amount of carbon at each time step, is not necessarily the best approach if current load shifts will impact the load profile in future time steps. This demonstrates that forecasting future load and generation patterns is important to obtain better solutions. Finally, the information needed to calculate the locational marginal carbon emissions is currently not made publicly available. Therefore, finding ways to infer and predict λ_{CO_2} at the data center nodes from publicly available data is needed in order to implement (λ_{CO_2}-shift) in practice.

CRediT authorship contribution statement

Julia Lindberg: Conceptualization, Methodology, Software, Investigation, Visualization, Writing – original draft. **Bernard C. Lesieutre:** Conceptualization, Methodology, Writing – review and editing, Supervision, Funding acquisition. **Line A. Roald:** Conceptualization, Methodology, Writing – review and editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bernard C. Lesieutre, Line Roald reports financial support was provided by NSF.

Acknowledgments

The authors would like to acknowledge funding from the U.S. National Science Foundation CRISP program under award 1832208, as well as funding from VMWare as part of the NSF/VMWare partnership on Next Generation of Sustainable Digital Infrastructure.

References

[1] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, Jonathan Koomey, Recalibrating global data center energy-use estimates, Science 367 (6481) (2020) 984–986.
[2] Pierre Delforge, Josh Whitney, Data Center Efficiency Assessment-Scaling Up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers, Natural Resources Defense Council, 2014.
[3] Yevgeniy Sverdlik, Analysts: There are Now More than 500 Hyperscale Data Centers in the World, Data Center Knowledge, Informa Tech Division of Informa PLC, 2017, https://bit.ly/3jx8lsz.
[4] Google, Google Environmental Report 2019, Technical report, Google, 2019, https://services.google.com/fh/files/misc/google_2019-environmental-report.pdf.
[5] Justine Calma, Amazon Boosts Climate Commitments and Greenhouse Gas Emissions, The Verge, 2020, https://bit.ly/2Sr4Vfc.
[6] Ana Radovanovic, Our Data Centers Now Work Harder When the Sun Shines and Wind Blows, Data Centers and Infrastructure, Google Official Blog, Google, 2020, https://bit.ly/3cXPAMu.
[7] Ana Radovanovic, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, Saurav Talukdar, Eric Mullen, Kendal Smith, Mari Ellen Cottman, Walfredo Cirne, Carbon-Aware Computing for Datacenters, 2021, arXiv Online available: https://arxiv.org/abs/2106.11750.
[8] Lancium, With 5th Patent, Lancium Powers Ahead in Fast-Ramping Data Center Innovation, https://lancium.com/lancium-smart-response-patent-press-release/.
[9] Fan Yang, Andrew A. Chien, ZCCloud: Exploring wasted green power for high-performance computing, in: IEEE Int Parallel and Distributed Processing Symposium, IPDPS, IEEE, 2016, pp. 1051–1060.
[10] Andrew A. Chien, Zero-carbon compute project, 2019, https://bit.ly/34m68dh.
[11] J Lindberg, Y. Abdennadher, J Chen, B C. Lesieutre, L Roald, A guide to reducing carbon emissions through data center geographical load shifting, in: ACM Int. Conf. on Future Energy Systems, 2021, pp. 430–436.
[12] Adam Wierman, Zhenhua Liu, Iris Liu, Hamed Mohsenian-Rad, Opportunities and challenges for data center demand response, in: Int. Green Comp. Conf., 2014, pp. 1–10.
[13] Zhenhua Liu, Adam Wierman, Yuan Chen, Benjamin Razon, Niangjun Chen, Data center demand response: avoiding the coincident peak via workload shifting and local generation, in: ACM Int. Conf. on Measurement and Modeling of Computer Systems, 2013.
[14] Zhenhua Liu, Iris Liu, Steven Low, Adam Wierman, Pricing data center demand response, in: ACM Int. Conf. on Measurement and Modeling of Computer Systems, 2014.
[15] Zhi Zhou, Fangming Liu, Zongpeng Li, Bilateral electricity trade between smart grids and green datacenters: Pricing models and performance evaluation, IEEE J. Sel. Areas Commun. 34 (12) (2016) 3993–4007.
[16] Zhi Zhou, Fangming Liu, Shutong Chen, Zongpeng Li, A truthful and efficient incentive mechanism for demand response in green datacenters, IEEE Trans. Parallel Distrib. Syst. 31 (1) (2020) 1–15.
[17] Zhi Zhou, Fangming Liu, Zongpeng Li, Hai Jin, When smart grid meets geo-distributed cloud: An auction approach to datacenter demand response, in: IEEE Conf. on Computer Communications, INFOCOM, 2015, pp. 2650–2658.
[18] Shutong Chen, Lei Jiao, Lin Wang, Fangming Liu, An online market mechanism for edge emergency demand response via cloudlet control, in: IEEE Conf. on Computer Communications, INFOCOM, 2019, pp. 2566–2574.
[19] Jie Li, Zhen Bao, Zuyi Li, Modeling demand response capability by internet data centers processing batch computing jobs, IEEE Trans. Smart Grid 6 (2) (2014) 737–747.
[20] Lei Rao, Xue Liu, M.D. Ilic, Jie Liu, Distributed coordination of internet data centers under multiregional electricity markets, Proc. IEEE 100 (1) (2011) 269–282.
[21] Hui Dou, Yong Qi, Wei Wei, Houbing Song, Carbon-aware electricity cost minimization for sustainable data centers, IEEE Trans. Sustain. Comput. 2 (2) (2017) 211–223.
[22] Lei Rao, Xue Liu, Le Xie, Wenyu Liu, Minimizing electricity cost: optimization of distributed internet data centers in a multi-electricity-market environment, in: 2010 Proc. IEEE INFOCOM, 2010, pp. 1–9.
[23] Weiqi Zhang, Line A. Roald, Andrew A. Chien, John R. Birge, Victor M. Zavala, Flexibility from networks of data centers: A market clearing formulation with virtual links, in: 21st Power Systems Computation Conference, 2020.
[24] Íñigo Goiri, William Katsak, Kien Le, Thu D Nguyen, Ricardo Bianchini, Parasol and greenswitch: Managing datacenters powered by renewable energy, ACM SIGPLAN Not. 48 (4) (2013) 51–64.
[25] Zhenhua Liu, Yuan Chen, Cullen Bash, Adam Wierman, Daniel Gmach, Zhikui Wang, Manish Marwah, Chris Hyser, Renewable and cooling aware workload management for sustainable data centers, in: 12th ACM SIGMETRICS/PERFORMANCE Joint Int. Conf. on Measurement and Modeling of Computer Systems, 2012, pp. 175–186.
[26] Kibaek Kim, Fan Yang, Victor M. Zavala, Andrew A. Chien, Data centers as dispatchable loads to harness stranded power, IEEE Trans. Sustain. Energy 8 (1) (2016) 208–218.
[27] Jiajia Zheng, Andrew A. Chien, Sangwon Suh, Mitigating curtailment and carbon emissions through load migration between data centers, Joule (2020).
[28] Z. Liu, M. Lin, A. Wierman, S. Low, L.L.H. Andrew, Greening geographical load balancing, IEEE/ACM Trans. Netw. 23 (2) (2015) 657–671.
[29] Andrew A. Chien, Richard Wolski, Fan Yang, The zero-carbon cloud: High-value, dispatchable demand for renewable power generators, Electr. J. 28 (8) (2015) 110–118.
[30] Fan Yang, Andrew A. Chien, Large-scale and extreme-scale computing with stranded green power: Opportunities and costs, IEEE Trans. Parallel Distrib. Syst. 29 (5) (2017) 1103–1116.
[31] Tomorrow, Data Driven Climate Action, Technical report, 2020, https://www.electricitymap.org.
[32] CAISO, Today's Outlook: Emissions, Technical report, 2021.
[33] CAISO, Managing Oversupply, Technical report, 2021.
[34] Kyle Siler-Evans, Ines Lima Azevedo, M. Granger Morgan, Marginal emissions factors for the US electricity system, Environ. Sci. Technol. 46 (9) (2012) 4742–4748.
[35] D.S. Callaway, M. Fowlie, G. McCormick, Location, location, location: The variable value of renewable energy and demand-side efficiency resources, J. Assoc. Environ. Resour. Econ. 5 (1) (2018) 39–75.
[36] David Oates, Kathleen Spees, Locational Marginal Emissions: A Force Multiplier for the Carbon Impact of Clean Energy Programs, Technical report, 2021.
[37] P. A. Ruiz, A. Rudkevich, Analysis of marginal carbon intensities in constrained power networks, in: 43rd Hawaii Int. Conf. on System Sciences, 2010.
[38] A. Rudkevich, P. A. Ruiz, R. C. Carroll, Locational carbon footprint and renewable portfolio policies: A theory and its implications for the eastern interconnection of the US, in: Hawaii Int Conf on System Sciences, HICSS, 2011.
[39] Richard Tabors, Aleksandr Rudkevich, Hua He, Ninad Kumthekar, Xindi Li, Robert Bland, Garret Quist, Joshua LaPenna, Methodology for calculation of the marginal emission rates from a complex cogeneration facility compared with that of the co-located NY ISO bus, in: Hawaii Int Conf on System Sciences, HICSS, 2021.
[40] Michelle M. Rogers, Yang Wang, Caisheng Wang, Shawn P. McElmurry, Carol J. Miller, Evaluation of a rapid LMP-based approach for calculating marginal unit emissions, Appl. Energy 111 (2013) 812–820.
[41] Y. Wang, C. Wang, C.J. Miller, S.P. McElmurry, S.S. Miller, M.M. Rogers, Locational marginal emissions: Analysis of pollutant emission reduction through spatial management of load distribution, Appl. Energy 119 (2014) 141–150.
[42] PJM, Marginal emission rates added to data miner tool, 2021, Online available: https://insidelines.pjm.com/marginal-emission-rates-added-to-data-miner-tool/.
[43] PJM, Marginal emissions rate – a primer, 2021, Online available: https://pjm.com/-/media/etools/data-miner-2/marginal-emissions-primer.ashx.
[44] PJM, 5 minute marginal emission rates, 2021, Online available: https://dataminer2.pjm.com/feed/fivemin_marginal_emissions/definition.
[45] J. Lindberg, B. C. Lesieutre, L. Roald, The environmental potential of hyper-scale data centers: Using locational marginal CO2 emissions to guide geographical load shifting, in: Hawaii Int Conf on System Sciences, HICSS, 2021, pp. 1–10.
[46] R.D. Christie, B.F. Wollenberg, I. Wangensteen, Transmission management in the deregulated environment, Proc. IEEE 88 (2) (2000) 170–195.
[47] E. Litvinov, Design and operation of the locational marginal prices-based electricity markets, IET Gen., Trans. Distrib. 4 (2010) 315–323(8).
[48] D. Bertsimas, J. Tsitsiklis, Introduction to Linear Optimization, 1997.
[49] C. Barrows, E. Preston, A. Staid, G. Stephen, J.-P. Watson, A. Bloom, A. Ehlen, J. Ikaheimo, J. Jorgenson, D. Krishnamurthy, J. Lau, B. McBennett, M. O'Connell, The IEEE reliability test system: A proposed 2019 update, IEEE Trans. Power Syst. 35 (1) (2020) 119–127.